The Influence of Computer-Mediated Feedback on Collaboration

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität München

vorgelegt von
Diplom-Medieninformatikerin
Sarah Tausch

München, den 09.08.2016

Erstgutachter: Prof. Dr. Heinrich Hußmann
Zweitgutachter: Prof. Dr. Carl Gutwin

Tag der mündlichen Prüfung: 19.10.2016

ABSTRACT

Collaboration is a crucial behavior of humans. People work together in a variety of contexts. For example, they wish to discuss and solve problems, to exchange knowledge or to be creative. However, group work is not necessarily successful when groups are left to themselves. The present thesis aims at improving collaborative work by means of technology. One technological solution to support small groups in being more effective is the use of *group mirrors*. These are systems that provide feedback to a group about specific aspects of their collaborative activities. An exemplary scenario is a small group that convenes a meeting to collect ideas on a certain topic. In one well-known realization of a group mirror, the speaking time of each group member is captured and displayed on a peripheral wall display, leading to an increased awareness of varying participation rates.

Previous research on group mirror systems mainly focused on providing quantitative feedback such as speaking times or speaking turns. This thesis focuses on the various ways of giving *qualitative* feedback during co-located group work. In order to cover the wide spectrum of possible collaboration setups but still stay focused, two representative tasks are evaluated in detail, which stand for extreme positions of the whole range of applications: collaborative creativity, a more open-ended task, and collaborative argumentation, a more structured task.

The contribution of this dissertation is twofold. Firstly, a design space for group mirrors is defined on a rather abstract and general level, and existing group mirrors are classified according to this design space. Secondly, for the two specific tasks, several aspects of this design space are systematically evaluated in studies using different prototypes, leading to suggestions for the design of group mirrors.

To evaluate the feasibility of group mirrors for collaborative creativity, we implemented four different prototypes that support *brainstorming* and the *Disney Method*, a creativity technique that makes use of different roles (i.e., *dreamer*, *realist* and *critic*). We used the amount of ideas as qualitative feedback that we showed to the group. To investigate the influence of different aspects of group mirrors on collaboration, we compared several display environments in form of visualizations on table or wall and public or private display settings. Finally, we addressed the problem of the competitive nature of existing group mirrors that can lead to frustration and social pressure by proposing more cooperative concepts.

Compared to collaborative creativity, argumentative debates represent a more structured task. An essential rule is to observe a particular structure of arguments. To support novices in learning how to use this structure, we built two prototypes. One system is composed of cylindrical light objects that facilitate peer feedback. We compared two versions of the light cylinders in a study in which the feedback providers either are anonymous or identifiable. A second system runs on smartphones and tablets and supports traditional debates as practiced in debate clubs by enabling a feedback loop between a jury member and the speaker.

All prototypes presented in the current thesis are classified according to the design space. A synthesis of the results is presented and suggestions for adequate support in various usage scenarios are derived. With this, we provide insights into the various effects of group mirrors on collaboration and intend to offer guidance for the design of technologically mediated feedback.

ZUSAMMENFASSUNG

Die Zusammenarbeit in Gruppen ist ein essenzielles menschliches Verhalten. Menschen diskutieren und lösen Probleme miteinander, tauschen untereinander Wissen aus oder sind gemeinsam kreativ. Gruppen arbeiten allerdings nicht unbedingt erfolgreich zusammen, wenn sie auf sich selbst gestellt sind. Das Ziel der vorliegenden Arbeit ist es, technologische Möglichkeiten vorzustellen, die die Zusammenarbeit innerhalb von Gruppen fördern können. Eine dieser Möglichkeiten kleinere Gruppen darin zu unterstützen effektiver zusammenzuarbeiten, sind sogenannte *Group Mirrors*. Das sind Systeme, die Feedback über das Verhalten in der Gruppe geben. Ein Szenario könnte sein, dass sich Personen treffen, um Ideen zu einem bestimmten Thema zu sammeln. Ein bekanntes Beispiel eines Group Mirrors zeigt auf einem peripheren Bildschirm an, wie viel jede Person gesprochen hat, was dazu führt, dass die Gruppenmitglieder diesem Aspekt mehr Aufmerksamkeit schenken.

Vorherige wissenschaftliche Arbeiten zu Group Mirror-Systemen verwendeten meist quantitatives Feedback, wie zum Beispiel die Anzeige von Redezeiten oder die Reihenfolge der Sprecher. Die hier vorgestellte Arbeit beschäftigt sich nun mit den verschiedenen Möglichkeiten, einer Gruppe *qualitatives* Feedback zu geben. Um ein möglichst großes Spektrum der vielen möglichen Arten von kommunikativen Gruppenprozessen zu erfassen, wurden zwei repräsentative Aufgaben gewählt, die zwei extreme Positionen abdecken, und zwar Kreativitätstechniken, die ein eher unbestimmtes Ergebnis haben, und Argumentationen, die einen strukturierteren Ansatz verfolgen. Dies soll im Wesentlichen durch zwei Beiträge erzielt werden. Zum einen wird ein eher abstrakter *Design Space* für die Gestaltung von Group Mirror-Systemen vorgestellt und bestehende Group Mirror-Systeme werden in diesen eingeordnet. Zum anderen werden verschiedene Aspekte dieses *Design Space* anhand der beiden zuvor vorgestellten Aufgaben systematisch evaluiert. Dazu wurden verschiedene Prototypen entwickelt, in Studien untersucht und Schlussfolgerungen für das Design von Group Mirror-Systemen abgeleitet.

Um die Eignung von Group Mirror-Systemen zur Unterstützung von kreativem Arbeiten in der Gruppe zu analysieren, entwickelten wir vier Prototypen, zwei davon für die Kreativitätstechnik *Brainstorming* und zwei für die *Disney Methode*, die auf der Verwendung verschiedener Rollen (*Träumer, Realist und Kritiker*) aufbaut. Qualitatives Feedback wurde der Gruppe dabei in Form der Anzahl der generierten Ideen gegeben. Um den Einfluss verschiedener Aspekte von Group Mirror-Systemen auf die Gruppenarbeit zu untersuchen, verglichen wir mehrere Display-Umgebungen miteinander. Insbesondere wurden Visualisierungen auf horizontalen oder vertikalen Oberflächen sowie der Einfluss von öffentlichen und privaten Display-Umgebungen untersucht. Zudem betrachteten wir das Problem, dass viele Group Mirror-Systeme einen eher kompetitiven Charakter haben, der zu Frustration und sozialem Druck führen kann. Daher entwickelten wir kooperative Konzepte und verglichen diese mit den traditionellen, eher kompetitiven Ansätzen.

In Bezug auf Kreativität haben argumentative Debatten eher eine strukturiertere Aufgabe. Eine der grundlegenden Regeln ist es, Argumente einer bestimmten Struktur nach aufzu-

bauen. Wir entwickelten zwei Prototypen, die dazu beitragen sollen, dass Anfänger diese Struktur verinnerlichen. In einem Ansatz werden Lichtzylinder verwendet, die den Gruppenmitgliedern ermöglichen sich gegenseitig Feedback zu geben. In einer Studie verglichen wir zwei Versuchsanordnungen, eine, in der derjenige, der das Feedback gibt, anonym bleibt und eine, in der diese Person identifizierbar ist. Ein weiteres System wurde umgesetzt, um traditionelle Debatten, wie sie in Debattierclubs praktiziert werden, zu unterstützen. Dabei wird eine Feedbackschleife zwischen der Jury und dem Sprecher ermöglicht, die dazu ein Smartphone und ein Tablet verwenden.

Alle genannten Prototypen werden schließlich in den *Design Space* eingeordnet und eine Synthese der vorgestellten Ergebnisse wird präsentiert. Vorschläge dafür, wie verschiedene Anwendungsszenarien unterstützt werden können, werden unterbreitet. Ziel dieser Arbeit ist es, damit Anhaltspunkte für die Gestaltung von technologisch vermitteltem Feedback zu geben.

DISCLAIMER

Personal Contribution Statement

In this thesis, I describe a number of projects that I executed in collaboration with several students and colleagues. Below, I declare my personal contribution to the different projects.

Chapter 5: Supporting Brainstorming
The GROUPGARDEN project is based on the bachelor thesis of Ismail Kosan (2013) and the project thesis of Andrey Raltchev (2013). I developed the initial idea of the project and set the frame for both theses. The work was then conducted in close collaboration with both students. In regular meetings, each step was jointly discussed. However, the key decisions (e.g., the study design, study execution and evaluation) were made by me. The work was published at the *Nordic Conference on Human-Computer Interaction (NordiCHI '14)* (Tausch et al., 2014) together with my co-authors Doris Hausen, Ismail Kosan, Andrey Raltchev and Heinrich Hußmann. The main part of the paper was written by me. I also worked in revisions based on the feedback from my co-authors and reviewers.

The second project is based on the bachelor thesis of Stephanie Ta (2014). I developed the initial idea of the project and set the frame for the thesis. The work was then conducted in close collaboration. In regular meetings, each step was jointly discussed. However, the key decisions (e.g., the study design, study execution and evaluation) were made by me. The work was published at the *CHI Conference on Human Factors in Computing Systems (CHI '16)* (Tausch et al., 2016) with Stephanie Ta and Heinrich Hußmann as co-authors. The main corpus of the paper was written by me. I also worked in revisions based on the feedback from the co-authors and reviewers.

The third project is based on the bachelor thesis of Katharina Sachmann (2014). I developed the initial idea of the project and set the frame for the thesis. The work was then conducted in close collaboration. In regular meetings, each step was jointly discussed. However, the key decisions (e.g., the study design, study execution and evaluation) were made by me.

Chapter 6: Supporting the Disney Method
The first project is based on the practical work of Fabius Steinberger (2013). We developed the initial idea of the project together. The work was then conducted in close collaboration. In regular meetings, each step was jointly discussed. I was the key decision maker for the study design. The study was executed together by both of us. The work was published at the *International Conference on Human-Computer Interaction (INTERACT '15)* (Tausch et al., 2015b) with Fabius Steinberger and Heinrich Hußmann as co-authors. The main corpus of the paper was written by me. I worked in revisions based on the feedback from the co-authors and reviewers.

The second project is based on the bachelor thesis of Fabian Nußberger (2014). I developed the initial idea of the project and set the frame for the thesis. The work was then conducted in close collaboration. In regular meetings, each step was jointly discussed. However, the key decisions (e.g., the study design, study execution and evaluation) were made by me. Part of the work was published under the extended abstracts category of the *CHI Conference on Human Factors in Computing Systems (CHI '14)* (Tausch et al., 2015a) with Fabian Nußberger and Heinrich Hußmann as co-authors. The main corpus of the paper was written by me. I also worked in revisions based on the feedback from the co-authors and reviewers.

Chapter 7: Supporting Collaborative Argumentation
The first project is based on the master thesis of Cornelia Reithmeier (2013). I set the goals and the frame for the thesis. The initial idea was developed by me. The concept of the prototype was then developed together. The whole work was conducted in close collaboration. In regular meetings, each step was jointly discussed. However, the key decisions (e.g., the study design, study execution and evaluation) were made by me.

The second project is based on the bachelor thesis of Andrea Attwenger (2015). The initial idea was developed by me and I set the frame for the thesis. The whole work was conducted in close collaboration. In regular meetings, each step was jointly discussed. However, the key decisions (e.g., the study design, study execution and evaluation) were made by me.

Chapter 8: Supporting Debates
The first project was conducted in close collaboration with Bernd Huber. The initial idea was developed by me. Then, all steps of the project were discussed together and we both had the same influence on decisions that were made. While I was more involved in the study design, Bernd Huber conducted the main part of the implementation of the prototype. Part of the work was published in the extended abstracts category of the *CHI Conference on Human Factors in Computing Systems (CHI '14)*, with Bernd Huber, Heinrich Hußmann and me as authors (Huber et al., 2014). Bernd Huber and I wrote the paper together with approximately equal shares. We both worked in feedback by Heinrich Hußmann and external reviewers.

The second project is based on the master thesis of Tobias Rindlbacher (2015). I developed the initial idea of the project and set the frame for the thesis. The work was then conducted in close collaboration, in the beginning also together with Bernd Huber. In regular meetings, each step was jointly discussed. However, the key decisions (e.g., the study design, study execution and evaluation) were made by me.

Grammatical Choices

As just described, I conducted several projects in close collaboration with students and colleagues. It would not have been possible to accomplish all of these projects without their help. Thus, I will use the scientific plural in these cases. For the work that I conducted alone, I will use the singular form.

ACKNOWLEDGMENTS

The work presented in this thesis would not have been possible without the support of many wonderful people. I want to express my deepest gratitude to everyone who has been there for me over the last years. The following mentions are certainly not exhaustive but I hope that everyone that I missed knows that I am very thankful for all the help and encouragement.

First and foremost, I would like to thank my supervisor Heinrich Hußmann for his great support in every undertaking during my whole time as a doctoral student. I would especially like to thank him for his guidance, his always excellent feedback and his patience. I also would like to thank Carl Gutwin for joining my committee as a second supervisor, for his time and effort and for sharing his external view on the topic.

I had the pleasure to work in an extremely supportive and friendly lab and many of my colleagues turned into close friends. I would like to thank Julie Wagner for always helping me to think out of the box, and for a lot of dancing, Alina Hang for having that special connection that made us wear partnerlook by chance so often, Tobias Seitz for the great experience to make music together and for forming a band with me, Hanna Schneider for talking about anything and everything with me and for sharing hotel rooms and festival tents, Emanuel von Zezschwitz for his positive vibes that can brighten every situation, Henri Palleis for being the best office mate that I could have imagined during the time I was writing my thesis, Simon Stusak for experiencing so many conferences, travels and festivals together, Axel Hösl for sharing his philosophical view on research, many interesting and a lot of funny conversations, Sebastian Löhmann for giving me such a warm welcome to the lab and for going to many concerts together, Daniel Buschek for helping me with my first CHI paper and for keeping me updated with all news about my favorite animal (donkeys), Bernhard Slawik for his contagious enthusiasm for technology, Doris Hausen for helping me with the writing of my first paper, Alexander De Luca for always discussing study designs and statistics with me, Malin Eiband for interesting talks during breakfast in the Yosemite national park, and Renate Häuslschmid for having always some interesting and amusing stories to tell. I furthermore want to thank my former and current colleagues Max Maurer, Florian Alt, Mariam Hassib, Bastian Pfleging, Mohamed Khamis, Maria Fysaraki, Daniel Ullrich and Alexander Wiethoff for many fruitful discussions.

My thanks also go to Andreas Butz for providing valuable feedback during several IDC presentations and for always encouraging me in my sportive endeavors. Furthermore, I want to thank Sheelagh Carpendale and the whole InnoVis Lab for offering such an inspiring environment for our joint project.

Many thanks also go to Franziska Schwamb for being caring, helping me with literally everything and for being interested in my research, and to Rainer Fink for helping with all sorts of technical issues that occurred during my studies.

It would not have been possible to accomplish this much work without the collaboration with several students who all contributed to this work in one or the other way. My special thanks go to Andrea Attwenger, Ida Buchwald, Simon Gurn, Bernd Huber, Ismail Kosan, Fabian Nußberger, Andrey Raltchev, Florian Reiner, Cornelia Reithmeier, Tobias Rindlbacher, Katharina Sachmann, Matthias Schulte-Uentrop, Fabius Steinberger, Stephanie Ta and Attila von Thermann. I also want to thank Michael Mirwald for his help in several studies and Jana Antosch-Bardohn for exchanging views on our research topics.

I especially want to thank Ingrid Allard-Tausch for proofreading my thesis and for always discussing my work with me when I needed someone to talk to. I furthermore want to thank Daniel Buschek, Malin Eiband, Alina Hang, Moritz Neugebauer, Henri Palleis, Bastian Pfleging, Hanna Schneider, Tobias Stockinger, Simon Tausch and Emanuel von Zezschwitz for proofreading parts of my thesis.

I furthermore would like to thank my friends. Special thanks go to Alice Thudt for the many discussion that we had about our research topics and for suffering setbacks and celebrating successes together.

Finally, I would like to express my deepest gratitude to my family for always being there for me. I am very grateful for the encouragement and support of my parents, Ingrid Allard-Tausch and Hubert Tausch, and my brother, Simon Tausch. Last but not least, I want to thank Moritz Neugebauer for going through good and bad times with me. Thank you for your love and patience.

TABLE OF CONTENTS

List of Figures

LIST OF TABLES

GLOSSARY

ARS	Audience Response System
BBS	Bulletin Board System
BCI	Brain Computer Interface
CCS	Classroom Communication System
CET	Cognitive Evaluation Theory
CMC	Computer Mediated Communication
CPM	Computerized Performance Monitoring
CSCL	Computer Supported Collaborative Learning
CSCW	Computer Supported Cooperative Work
EBS	Electronic Brainstorming System
GSS	Group Support System
HCI	Human Computer Interaction
PFP	Pay For Performance
PoI	Points of Information
SA	Situation Awareness
SDT	Self-Determination Theory
TEC	Theory of Explanatory Coherence
TTCT	Torrance Test of Creative Thinking
WA	Workspace Awareness

1 Introduction

Cooperation and collaboration are essential skills in our daily life. We work together in a variety of different situations and contexts, be it to solve problems, to exchange knowledge, to learn from each other or to be creative together. Research agrees about this prominent role of collaboration. Barnard (1938), for instance, claims that cooperation is the most effective way to overcome the biological limitations of humans. Through the introduction of different technological solutions, new opportunities for collaboration have opened up. Enabling easy communication over distance, for example, has led to the possibility to collaborate with people from all over the world. Supporting remote collaboration is an important area of research. Still, a lot of collaborative activities continue to be accomplished in co-located sessions. Though co-located collaboration might seem less challenging to accomplish than remote situations, it is not necessarily effective when groups remain without support. Some people might dominate the group work, while others might restrain themselves from taking part actively, since they fear being evaluated by others, or since they feel that their input to the group work is not valuable (Diehl and Stroebe, 1987).

Group mirrors are one possibility to support co-located collaboration through technology and to address these problems. These systems reflect certain aspects of collaborative activities to a group (Jermann et al., 2001). To illustrate the concept of group mirrors, I allege an example of several existing group mirrors from related research. In a collaborative situation, microphones capture speaking times of the individual group members. This information is then visualized on a peripheral display by representing the speaking times of the group members with a bar chart, for example.

The deployment of group mirrors entails a number of positive effects on group work, suggesting that these can help to structure and guide collaboration. This is particularly valuable when there is no moderator or teacher present. Group mirrors, in contrast to verbal feedback, make it possible to provide feedback in a subtle and unobtrusive way without interrupting the group work. The possibility to give feedback in real-time, moreover, leads to a shorter feedback loop. Previous research reports about effects such as more balanced participation and an increased awareness of group processes (see Chapter 2).

However, there is also a number of problems and unsolved questions about group mirrors. They might, for instance, distract from the main task and make group members feel under pressure (see e.g. Bachour et al., 2008; Schiavo et al., 2014). Results from previous research indicate that the concept and design of group mirrors have a large influence on its success. For instance, Streng et al. (2009) could show that different visualizations affect self-regulation of the group and on the acceptance of the group mirror.

1.1 Problem Statement and Research Objectives

Designing studies with the goal to evaluate the effects of group mirrors on collaboration is a challenging task. There are two main reasons for that. First, it is an already complex endeavor to understand collaborative processes. This complexity even increases by introducing technological support to such situations. While it is already demanding to comprehend the effects of technology on individuals, revealing the effects on groups is even more difficult, since group dynamics have a large impact on the behavior of group members. The second main reason is the high amount of possible confounding variables that are inherent to technologically supported collaboration. These include aspects about the compilation of the group (e.g., group size or familiarity of the group members), as well as factors of the group mirror system itself (e.g., type of visualization, placement of the group mirror or amount of guidance). One main goal of this thesis is to offer a systematic description of these possible variables. We therefore present a design space for group mirrors.

The second main objective of this thesis is to investigate several of the factors of the design space in more detail. A number of previous studies on group mirrors already indicated that it is important to consider the concept and design of group mirrors to increase positive and to decrease negative effects in different usage scenarios. However, a large number of the different characteristics of group mirrors have not been investigated systematically yet. To this end, we analyzed several factors such as the display environment or the concept underlying the visualizations and provide suggestions on how these aspects can be used in different scenarios to improve group work. In addition, we shifted our focus, compared to previous research, to more qualitative feedback, which was collected using methods such as peer feedback. This approach aims at broadening the insights of different types of group mirrors. Moreover, we picked two specific use cases for this thesis, creativity and argumentation. This allows to cover a possibly broad spectrum of tasks that reach from more open-ended and unstructured (i.e., creativity) to more determinate and structured (i.e., argumentation).

1.2 Main Contributions

With this thesis, I aim to contribute to the research on group mirrors in three ways. First, I present a design space for group mirrors. Second, a number of different factors of the design space are investigated in comparative studies. Third, two use cases of group mirrors (creativity and argumentation) are explored.

1.2.1 A Design Space for Group Mirrors

The design space presented in this thesis offers a structured approach to classify group mirrors according to a number of factors. These factors include the type of information that the group mirror displays, the type of visualization, the level of aggregation of the mirrored information, the placement and privacy of the feedback or the amount of guidance that a system provides. This design space can support designers of group mirrors to consider possible design choices in a structured way. Moreover, it may be of use for planning studies, since it offers information about the different variables of group mirrors. Finally, a classification of existing group mirrors according to this design space reveals areas that still need further evaluation.

1.2.2 Insights on Different Aspects of the Design Space

We investigated a number of factors of the design space that we assess as important aspects, both in the context of creativity and argumentation. In two studies, we analyzed different display environments. The first of these studies compared two versions of public displays in form of a table and a wall display with each other. The second study examined public and private display environments. With another prototype, we investigated the impact of more cooperative and more competitive visualization on performance and acceptance. In two studies, we evaluated the effect of peer feedback and compared anonymous feedback (i.e., the provider of the feedback stays anonymous) with identifiable feedback (i.e., the feedback provider is known to the feedback receiver). Finally, we compared feedback with varying complexity in the context of debates. These studies revealed a number of insights, for instance, that an increase of performance is often accompanied with a decrease of acceptance of the group mirrors.

1.2.3 An Exploration of Group Mirrors for Creativity and Argumentation

We explored two specific use cases for group mirrors that represent the extremes of a continuum from more open-ended to more determinate tasks. More open-ended tasks are represented by collaborative creativity techniques. Here, we investigated group mirrors designed for brainstorming and the Disney Method, a role-based creativity technique. More determinate tasks are represented by argumentation, which is a more structured task than collaborative creativity. In particular, we evaluated group mirrors for collaborative argumentation and for argumentative debates. In this way, we provide insights on the design of group mirrors for these tasks.

1.3 Thesis Overview

This thesis is structured in nine chapters which are in turn structured in five main parts (see Figure 1.1). In the first part, I provide an overview of existing research. In the second part, I outline a design space for group mirrors. The third part addresses group mirrors for creative tasks, while the fourth part deals with the support of argumentation. Finally, the fifth part draws a conclusion on research on group mirrors.

Chapter 1: Introduction The first chapter motivates the topic of group mirrors. It further outlines the research objectives and the main contributions.

Part I: Setting the Stage for Group Mirrors

Chapter 2: Feedback In the second chapter, I provide an overview of related research on feedback. After a general introduction that discusses feedback theories as well as the various characteristics and different definitions of feedback, I turn to technologically mediated feedback in collaborative environments. Finally, awareness systems are discussed.

Chapter 3: Application Areas The third chapter presents related research on the two application areas that were chosen for this thesis. First, I address collaborative creativity, including brainstorming and the Disney Method as two creativity techniques and talk about computer support for collaborative creativity and issues with the evaluation of creativity. Moreover, I discuss effects of feedback on creativity. The second part of this chapter presents related research on argumentation. The two use cases, collaborative argumentation and debates, are defined. I then present a number of existing systems aiming at supporting argumentation, discuss issues with the evaluation of arguments and summarize previous findings on the effects of feedback on argumentation.

Part II: A Design Space for Group Mirrors

Chapter 4: Design Space The fourth chapter discusses a design space for group mirrors. As part of this, previous classifications are summarized. Along with the factors of the CSCW design space, which have already been described in related research, I outline the design space for group mirrors and classify existing group mirrors according to this design space.

Part III: Supporting Creativity with Group Mirrors

Chapter 5: Supporting Brainstorming In the fifth chapter, I report about three projects aiming at supporting brainstorming. The first one, GROUPGARDEN, is a visualization that represents a group during brainstorming using metaphors derived from nature (e.g., flowers, tree). I discuss two studies, one evaluating the general feasibility of group mirrors for brainstorming and another one investigating the differences of group mirrors on tables and walls.

Chapter 6: Supporting the Disney Method In chapter six, I describe two prototypes that aim at supporting the Disney Method, a role based-creativity technique. Moreover, I discuss two studies, one field study and one laboratory study that we used to evaluate the two systems.

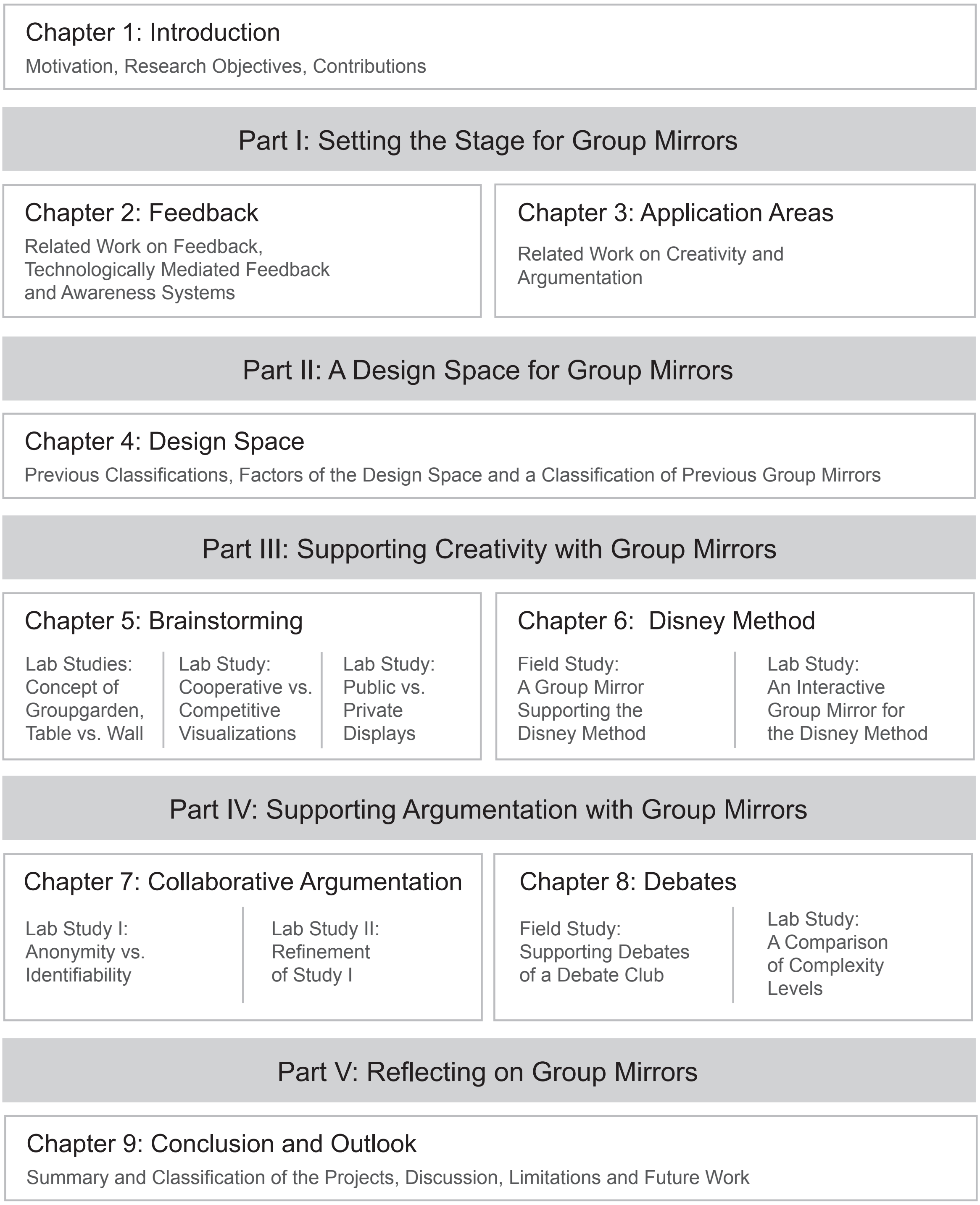


Figure 1.1: Thesis overview.

Part IV: Supporting Argumentation with Group Mirrors

Chapter 7: Supporting Collaborative Argumentation In the seventh chapter, the influence of technologically mediated peer feedback on collaborative argumentation is exploited. In the course of this chapter, I present a prototype consisting of interactive light cylinders that enables group members to provide feedback to each other. Two laboratory studies with slightly different study setups investigate the concept in general and specifically the influence of feedback that is provided anonymously in contrast to identifiable feedback.

Chapter 8: Supporting Debates Chapter eight discusses two prototypes designed for debates. First, I present a field study that aims at collecting early feedback from experts and novices. Then, I report about a more structured study that compares different complexity levels of feedback.

Part V: Reflecting on Group Mirrors

Chapter 9: Summary Chapter nine provides a summary by classifying the presented projects according to the design space. Moreover, I outline the main findings. I conclude my thesis by discussing the limitations of this work related to the study designs, the generalizability of the results, possible negative effects of group mirrors and the term *group mirror* that we used throughout the thesis. Finally, I provide suggestions for future work on group mirrors by naming possible use cases and by pointing out factors of the design space that to date have been widely neglected. I also discuss possible follow-up studies along with more general directions for future research.

I

SETTING THE STAGE FOR GROUP MIRRORS

2
Feedback

Feedback is powerful if it is done well. In decades of research the different aspects of feedback were evaluated to understand their effects better. In this section, I will give a brief overview of the related research on feedback and I will provide a definition, which specifies the meaning of the term in the scope of this thesis. I will then present related research on technologically mediated feedback in more detail.

2.1 A Brief Introduction to Feedback

The influence of feedback has been investigated in different contexts. The roots of research on feedback in the learning context reach back to the theory of the behavioristic law of effect by Thorndike (1913), in which feedback appears as "reinforcement" or "punishment". However, this theory is inconsistent with empirical results of several studies that show that feedback is not automatically effective. For reviews of this issue see for instance Annett (1969), Adams (1978) or the meta-analysis and summary of Kluger and DeNisi (1996). Since then, one major motive of research on feedback was to understand, under which circumstances feedback is valuable and helpful, not only in the context of learning. In the following, general theories on feedback are briefly presented, characteristics of feedback are outlined and a definition is provided.

2.1.1 Feedback Theories

A number of hypotheses on feedback were derived from theories that include feedback as a theoretical component (Kluger and DeNisi, 1996). The **control theory** of Carver and Scheier (1981) for instance is related to the approach of cybernetics. Wiener (1948) defines cybernetics as *"the science of communication and control"*, Carver and Scheier (1998) as *"the science of feedback processes; feedback processes involve the control or regulation of certain values within a system."* The control theory describes a negative feedback loop (see

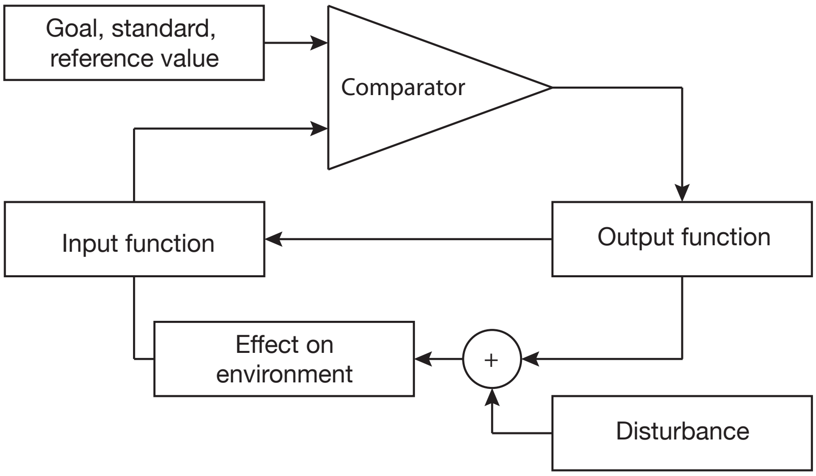


Figure 2.1: Feedback loop. The schematic depiction of a feedback loop by Carver and Scheier (1998)

Figure 2.1) that Carver and Scheier (1998) illustrate with the example of a thermostat: A thermostat measures the temperature of the air in a room and compares it to the desired temperature that is set in the thermostat. Dependent on the result of this comparison, the thermostat turns a heater on or off. By that, the thermostat tries to eliminate any discrepancies between the actual and the desired state. The example of the thermostat can be mapped on the schematic depiction of the feedback loop. The input function is denoted with the sensed temperature, the standard is the value that is set in the thermostat as the desired temperature. The comparator tries to identify discrepancies between the standard and the input function, which is the sensed temperature in the example. The output function is the action of turning on the heater. In the example, disturbances are external variables such as sunlight or the number of people in the room.

In contrast to the control theory, where the main assumption is that people want to eliminate a discrepancy between standard and current state, the **goal setting theory** by Latham and Locke (1991) assumes that people rather want to achieve a goal. The theory states that goals motivate action, and that feedback and goals combined are effective for motivating high performance. However, both goal setting without feedback and feedback without goal setting are not very effective.

Control theory and goal setting theory are, although important, only two examples among many. Kluger and DeNisi (1996) conducted a meta-analysis of 131 studies. Their analysis confirms that feedback does not always improve performance. Based on their findings, they developed the **feedback intervention theory**, which includes three levels of control: task-learning processes, task-motivation processes and meta-task processes that involve the self. These processes are ordered in a hierarchical way with task-learning processes at the bottom and meta-task processes at the top. With task-learning processes, Kluger and DeNisi (1996) refer to processes that involve the task-details, while task-motivation processes are mainly focused on the focal task. Meta-task processes include for instance attention to the self. One of the authors' conclusions is that feedback is specifically effective when it is provided on a familiar task and when it is directed to the task level rather than to the self level. Furthermore, their results show that feedback is more effective when goals are challenging and defined precisely and when feedback is given on correct responses instead of incorrect ones.

The conceptual analysis of Hattie and Timperley (2007) provides additional evidence that feedback is only successful under certain circumstances. They derive a model of feedback from their conceptual analysis of feedback in the context of learning. They claim that effective feedback has to answer the questions (1) *Where am I going?*, (2) *How am I going?* and (3) *Where to next?* The first question focuses on the goals (corresponding to the notion of *feed up*), the second on the progress toward a goal (corresponding to *feed back*) and the third on the activities (corresponding to *feed forward*). These questions work on four levels: The *task level*, the *process level*, the *self-regulation level* and the *self level*. Feedback on the *task level* is the most common type of feedback and describes feedback about the correctness of a task. As too much feedback on the *task level* can discourage students, this type of feedback is most effective when it leads students from the task to the *process level*. Feedback on *process level* focuses on the processes that underlie a certain task. An example for this level is strategies for error detection. Feedback on this level helps most when it provides learners with directions for searching and strategizing. *Self-regulation* includes for instance capabilities of self-assessment, attributions of success or failure or the way how people seek help. Feedback on the *self-regulation level* can engage students to invest further effort into the task and lead to better self-efficiency. Feedback at the *self level* includes positive or negative evaluations but with little task-related information. This type of feedback is rarely effective.

Several aspects of these feedback theories serve as a basis for research on group mirrors. On the one hand, these theories and studies are valuable because they provide a strong foundation for the design of group mirrors. On the other hand, further studies will be necessary to learn, if and how group mirrors differ from these theories or how they complement them. In the next section, characteristics of feedback are summarized that are part of several feedback theories and that are especially important in the context of group mirrors. Afterwards, definitions of feedback are summarized and a definition is provided that is suitable in the scope of the current thesis.

2.1.2 Characteristics of Feedback

Later in this thesis, I will present a design space for group mirrors. As some of the characteristics of group mirrors are related to the different forms and types of feedback in general, I will give a brief overview of the classifications linked to feedback. There exist variant other classifications of the different types of feedback in different contexts, e.g., by Lyster and Ranta (1997), Hattie and Timperley (2007) or Brookhart (2008). I will summarize some of the characteristics that are especially important in the context of this thesis.

Feedback Valence As noted before, feedback can either be positive or negative. Several studies support the conclusion that both types of feedback can be beneficial (see e.g. Hattie and Timperley, 2007; Kluger and DeNisi, 1996; Van-Dijk and Kluger, 2004) argue that the effectiveness of positive and negative feedback is dependent on the levels that it is aimed at. On the *task level*, corrective feedback is most effective. Furthermore, negative feedback accompanied with corrective information is more powerful than without this information (see e.g. Weiner, 1974), especially when attention is paid to a proper presentation of the feedback (Howie et al., 2000). Referring to the *self-regulation level*, Van-Dijk and Kluger (2001) were able to show that it makes a difference if people "want to do" or "have to do" a task. In the first case, positive feedback increases motivation, in the second case it leads to a decrease of motivation. This indicates that the attitude towards a certain task affects the effectiveness of feedback on the *self-regulation level*. Swann et al. (1988) additionally show different effects of feedback dependent on the self-efficiency of students. Referring to the *self level*, several studies provide evidence that negative feedback is more effective than positive when given on this level (see e.g. Brockner, 1979; Hattie, 1992; Kinch, 1963, 1968).

Comparison In learning sciences, there exists the differentiation between norm-referenced tests and criterion-referenced tests (see e.g. Block, 1971; Glaser, 1963; Glaser and Nitko, 1970; Hambleton and Novick, 1973). Criterion-referencing tests are *"deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards"* (Glaser and Nitko, 1970), while norm-referencing tests compare *"information about the capability of a student compared with the capability of other students"* (Glaser, 1963). This classification can be transferred to the context of feedback, that can also be either of *norm-referenced* or *criterion-referenced* nature. Studies revealed a number of advantages of criterion-referenced feedback in comparison to norm-referenced feedback in the context of learning (Crooks, 1988, Wilburn, 1983).

Timing of Feedback The aspect of timing addresses when feedback is provided (immediate or delayed) and how often it is given. Again, Hattie and Timperley (2007) discuss this aspect in the light of the different levels they identified. They summarize the results of Kulik and Kulik (1988), who report about advantages of delayed feedback on *task level* and immediate feedback on *process level*. Clariana et al. (2000) investigate the correlation between task difficulty and timing of feedback and suggest delayed feedback for more complex items and immediate feedback for easier items.

Function Another aspect is the *function* of the feedback, meaning that it can either be descriptive or evaluative, a categorization discussed by Tunstall and Gsipps (1996), for instance. Descriptive feedback includes an explanation with the goal to tell a student what and how to improve. Evaluative feedback in contrast tells a student how well he or she performed, for example, by giving grades, check marks or evaluative comments. Tunstall and Gsipps (1996) argue that, at least in the classroom, the combination of both descriptive and evaluative feedback is most powerful.

There is also a number of other characteristics of feedback that I will not discuss in detail here. It is nevertheless worth to mention them, as these also are factors that play a role for the design space of computer-mediated feedback systems. Some of them are described in more detail in Chapter 4. Brookhart (2008) mentions the **amount of feedback**, the **feedback**

mode (this factor is related to the aspect of "feedback modality" in the design space for group mirrors) and **audience** (referred to as "feedback receiver" in the design space). Other factors of the feedback content mentioned by Brookhart (2008) are **clarity** of the feedback to the student, **specificity** and **tone**.

2.1.3 Definitions of Feedback

Finally, I will discuss different definitions of feedback. Quite generally, *"feedback is information received by an individual about his or her past behavior"* (Annett, 1969). Kluger and DeNisi (1996) define feedback as *"actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one's task performance."* Hattie and Timperley (2007) define feedback in the context of learning: *"Feedback is conceptualized as information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding. A teacher or parent can provide corrective information, a peer can provide an alternative strategy, a book can provide information to clarify ideas, a parent can provide encouragement, and a learner can look up the answer to evaluate the correctness of a response. Feedback thus is a 'consequence' of performance."*

In the scope of the current thesis, feedback is provided to a group of people involved in a collaborative task. The goal of this feedback is to help group members to learn how to conduct a specific task rather than to assist the group with this feedback on a regular basis. Furthermore, the feedback is mediated through technology. This implies that a system might gather information and present the feedback, but also that a teacher or facilitator provides this feedback, however not orally but with the help of technology.

2.2 Technologically Mediated Feedback in Collaborative Environments

As the short summary of theories and aspects of feedback showed, the effectiveness of feedback depends highly on how the feedback is provided and in which context. With the increasing prevalence of technology, computer-mediated feedback became more and more common. Earley (1988) conducted one of the first studies investigating the effect of computer-based performance feedback. Results of that study show that feedback increases performance when it is received from a computer compared to feedback from a supervisor.

Another systematic approach of understanding the differences between person- and computer-mediated feedback was presented by Kluger and Adler (1993). They confirm results of Karabenick and Knapp (1988), who were able to show that people rather seek feedback provided from a computer than from a person. Kluger and Adler (1993) however assume that these results are mainly true for people seeking objective performance feedback, while this might not be true for seeking feedback concerning impression management motives (Ashford and Tsui, 1991), in which case people seek an opportunity to impress others.

Alder and Ambrose (2005) investigated the effects of person- and computer-mediated feedback in the context of computerized performance monitoring (CPM) systems, defined as *"any method of collecting, storing, analyzing, and reporting individual or group actions or performance on the job"* (Nebeker and Tatum, 1993). Their results indicate that face-to-face feedback was perceived fairer than computer-mediated feedback. In summary, results show that computer-mediated feedback seems to be beneficial under certain circumstances, for example, when it is provided on performance related aspects.

These results provide evidence that technologically mediated feedback has certain advantages in contrast to face-to-face feedback, however, it depends on the task and circumstances how effective the feedback is. In the following, I will first provide definitions for the concept of *group mirrors*, systems that provide technologically mediated feedback in collaborative situations. I will then give an overview on existing prototypes and results of several studies conducted with these systems. Afterwards, systems that are used to provide feedback in lectures are outlined, as several of these tools served as an inspiration for the group mirrors presented in this thesis.

2.2.1 Definitions of Group Mirrors

This thesis addresses the influence of computer-mediated feedback on group processes. Therefore, I will provide a more detailed analysis of technologically mediated feedback in the context of collaborative activities. According to several other researchers, we call systems that provide this feedback *group mirrors*. Jermann et al. (2001) uses the term *mirroring tools*, other authors also call these kind of tools *group mirrors* (e.g. Streng et al., 2009) or *social mirrors* (e.g. Karahalios, 2009). In the following, we review and extend existing definitions and provide an overview of the research on *group mirrors*.

Jermann et al. (2001) distinguish between *mirroring tools*, *meta-cognitive tools* and *guiding systems* (see Figure 2.2). All of these tools have in common that they initially collect data about collaborative processes. *Mirroring tools* use the aggregated data to mirror the current state to the group. *"Mirroring tools automatically collect and aggregate data about the student's interaction (...), and reflect this information back to the user (...)"* (Jermann et al., 2001). *Metacognitive tools* go one step further and compare the current state to the desired state and show this to the group members. By that, these tools stimulate metacognitive processes, as people are encouraged to think about possible discrepancies between these states. If there are discrepancies between the current and the desired state, there is also the option to provide advice and guidance on how this gap can be resolved by *guiding systems*. It is worth to mention that Jermann et al. (2001) make these distinctions in the context of computer-supported collaborative learning. The collected data in their examples is interaction data, captured for learners interacting with a learning tool. However, their definition is also valid for group mirrors outside the context of learning.

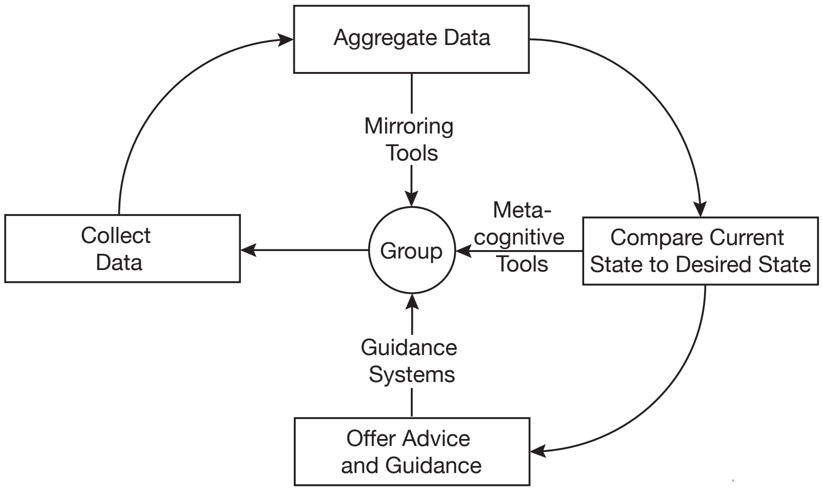


Figure 2.2: Definition of feedback systems. *Mirroring*, *meta-cognitive* and *guiding systems* based on Jermann et al. (2001) and Streng et al. (2009)

Karahalios (2009) describes three qualities of *group mirrors*. She calls these systems *social mirrors*. She defines that these tools allow people to see information about themselves in the context of information about other group members, while everyone sees the same visualization. The visualizations themselves are subtle and appear near real-time and groups can interact with the group mirror, e.g., through real-time experimentation, replay, annotation or reconfiguration.

We use a similar definition as Jermann et al. (2001), however, with one alteration. In our definition, the data is not necessarily collected automatically. We include systems in the definition of *group mirrors*, in which humans collect the data that is then reflected to the group. This allows *group mirrors* to show qualitative information, something that is difficult to achieve when data is collected automatically. In our definition, *group mirrors* can therefore also provide **computer-mediated** feedback in contrast to the previous definitions that only encompassed **computer-generated** feedback.

2.2.2 Group Mirrors for Co-located Collaboration

Existing group mirrors are mostly designed to support collaborative activities, such as supporting problem solving or knowledge sharing tasks. However, other use cases are possible, too. Ogawa et al. (2012) present a system called TABLE TALK ENHANCER, which is similar to some of the systems described in this section, but is designed specifically for mealtime communication, which is more a communicative than a collaborative situation. In the following, group mirrors focusing on collaborative tasks are outlined in more detail.

DiMicco et al. (2004) created one of the first group mirror systems designed to facilitate co-located group work. Based on a first prototype (see DiMicco, 2004; DiMicco and Bender, 2004) that displays key words of a group's conversation, the authors focused on a subtler and

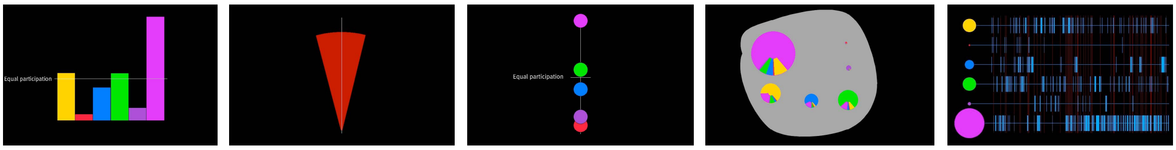


Figure 2.3: Second Messenger. This system of DiMicco and Hollenbach (2006) includes several visualizations that display speaking times: Histogram (left), Fan (second from left), Bouncing Balls (middle), Group Circle (second from right) and a Timeline (right).

less information-rich interface. The main goal of their ambient visualization is to facilitate higher quality decision making by helping groups to include more diverse viewpoints in a discussion. Their system, called Second Messenger, shows speaking times and turns on a peripheral wall display. Each group member wears a microphone that captures who is speaking for how long. This information is then visualized in form of a histogram. The individual bars of the bar chart are colored differently and numbered in ascending order. As one of the authors' goals is to imply a standard of the desirable group behavior, the height of the bars is labeled with "under", "participating" and "over". Each participant is represented by one bar. To help participants remember their representation, the numbers are placed in front of them on a table. At the top of the display, colored dots denote who has spoken in the last 30 seconds.

In their behavioral study (see also DiMicco et al., 2007) the authors compared groups involved in an information sharing task using this display with a baseline without group mirror visualization. They could show that "over-participators" decreased their participation with the aid of the group mirror. In general, participants thought that they contributed more than it was actually the case. The authors assume that this is one reason why "under-participators" did not increase their participation.

In a newer version of the Second Messenger (see Figure 2.3), DiMicco and Hollenbach (2006) add other visualizations such as a Fan, which in contrast to the histogram version summarizes speaking times over all participants and indicates an overall deviation from the optimal state. In the Bouncing Balls visualization, group members are represented by colored balls, which move to the top or bottom while the ball's vertical position mirrors the amount of participation. This visualization exists in an identifiable version in which individual group members can be recognized by different colors and an anonymous version in which all balls are monochrome. The Group Circle visualization represents group members with circles that grow and shrink dependent on participation levels. Overlapping speech can be displayed by clicking on a circle that then turns into a pie chart. The pie slices show, who spoke simultaneously with the respective person. Finally, the Timeline visualization is a version envisaged as a replay of the discussion. On the left, circles represent participants similar to the Group Circle visualization. Next to the circles, speaking times are displayed with vertical bars.

Several studies with these visualizations indicate that the Second Messenger can serve as an alternative for hand-coding group behavior by enabling an automated evaluation of

Figure 2.4: Related work on group mirrors. Top row from left to right: group mirror by Terken and Sturm (2010), the CONVERSATION CLOCK by Bergstrom (2006) and the REFLECT table by Bachour et al. (2010). Lower row: The MEETING MEDIATOR by Kim et al. (2008), the metaphoric group mirror by Streng et al. (2009) and the system by Schiavo et al. (2014).

group dynamics (DiMicco et al., 2006). DiMicco and Bender (2007) and DiMicco et al. (2007) also compared different combinations of lab and natural group settings and different visualizations shown in real-time and as a replay. They conclude that redundant feedback was harmful for the perception of feedback and that group members should be trained on the social purpose of the visualization. Moreover, participants found detailed feedback shown as a replay as most informative and useful.

Sturm et al. (2005), Kulyk et al. (2006) and Terken and Sturm (2010) present a system that additionally to feedback on speaking times includes information about gaze behavior of the group members (see Figure 2.4, top left). Each group member sees three circles that are projected on the table in front of them: one showing the speaking time, one showing the amount of visual attention received while speaking and one reflecting the attention received as a listener from other speakers. The current speaker additionally is notified of the current speaking time and the attention through lighter-colored rings surrounding the corresponding circle. Results of a study using a Wizard of Oz approach provide preliminary evidence that overall, speaking times increased, participants were more satisfied with group processes and visual attention was shared amongst listeners more evenly.

In a larger study, Sturm et al. (2007) used a fully functional system instead of the Wizard of Oz approach. Results confirm that groups participated in a more balanced way with feedback than without. Compared to the results of DiMicco and Bender (2007), not only "over-participators" reduced their speaking times, but also "under-participators" spoke more. As in the first study, a tendency of more evenly distributed visual focus of attention could be observed. However, in contrast to the conclusions of DiMicco and Bender (2007), the authors report that participants had concerns about the system distracting them from the discussion.

Bergstrom and Karahalios (2007a) present a similar approach to the SECOND MESSENGER (see also Karahalios and Bergstrom, 2006). Their system CONVERSATION CLOCK uses a visualization in form of concentric rings to display speaking times and turns (see Figure 2.4, top middle). In contrast to the SECOND MESSENGER, the visualization additionally shows the loudness of speech in form of the length of rectangular bars. These are colored differently for each group member and show up on a circular timeline when the corresponding person is speaking. Silent phases are indicated by small dots, overlapping speech by multicolored bars. A history of the past conversation is shown in form of concentric rings that move to the center when a new circle is completed.

In a study, Bergstrom and Karahalios (2007c) shed light on the question of how people perceive themselves and others with and without feedback through a group mirror visualization. They observed that participants were particularly interested in their own interactions, but that they are also more aware of other group members' interactions. Similarly to the results of the studies of DiMicco and Bender (2007), the presence of the CONVERSATION CLOCK prompted above average speakers to decrease the length of their contributions. Below average speakers however increased the number of speaking turns. While DiMicco and Bender (2007) conclude that the SECOND MESSENGER did not seriously distract group members or made them socially uncomfortable, Bergstrom and Karahalios (2007c) report that several participants perceived the visualization as distracting and tended to shift their gaze from the other participants to the group mirror on the table.

Building on this line of research, Bergstrom and Karahalios (2007b) present a tool called CONVERSATION VOTES that enables group members to cast anonymous votes on discussions. The visualization is similar to the CONVERSATION CLOCK, as it displays speaking times and turns using colored bars. In contrast to the CONVERSATION CLOCK, the length of the bars does not reflect the amplitude, but displays the amount of votes that were casted at a particular moment of the conversation. Additionally, dots at the end of the bars mark the exact time of a vote. Votes are made anonymously via physical buttons. While in a first version, group members could give positive feedback leading to larger bars as well as negative feedback leading to smaller bars, Bergstrom and Karahalios (2009) removed the negative feedback button in the iterative design process since the decreasing bars irritated participants.

In a study, Bergstrom and Karahalios (2009) point out that opening a back-channel during a discussion increases the awareness of own and others' contributions and, similarly to the before mentioned systems, facilitates balanced participation. Especially active and unsatisfied voters participate more and thereby strive for more diverse opinions to be included in the discussion.

To investigate the influence of visual feedback on group processes, Bergstrom and Karahalios (2012) conducted a study in which they purposefully distorted the feedback of the CONVERSATION CLOCK. Initially, they considered four distortion strategies: raising the speed by altering the sampling rate, increasing the amplitude indicating louder speech, decreasing the brightness of everyone apart from one group member, and color replacement, meaning that the speech is assigned to someone else. Building on a prestudy, the authors decided to use ‘speed’ as distortion strategy for their main study. They were able to show that people trust the visualization more than their own perception and reacted to the distorted group mirror similarly as to the undistorted one. This indicates that accuracy of the visualization has a minor impact on group processes and leaves reasons for discussions and evaluations about which factors actually have an impact on groups.

A system designed to balance participation in the learning context is the REFLECT table by Bachour et al. (2008) (see Figure 2.4, top right). The table shows colored LEDs that represent the amount of speech of the group members. In a user study, Bachour et al. (2010) compare their system with a baseline. In contrast to the before mentioned studies, the baseline is not a situation without any feedback, but a condition in which the LEDs represent the time spent on certain topics. This leads to a comparison of a speaker-based condition to a topic-based condition. Results show that with speaker-based feedback, participation levels are more balanced, however, only for group members who believed that balanced participation is important. Interestingly, for extreme participants (extreme “over-” or “under-participators”), topic-based feedback even had a reverse effect, meaning that “over-participators” tended to contribute more and “under-participators” less. These results add to the assumption of Bergstrom and Karahalios (2012), who claimed that the accuracy of feedback has a minor impact on group processes. The results of Bachour et al. (2010), however, suggest that at least not any kind of visualization of group related information can lead to more balanced participation.

The MEETING MEDIATOR, developed by Kim et al. (2008) and Kim and Pentland (2009) captures information about the group and group members via so-called “sociometric badges” (see Figure 2.4, bottom left). These include social signals such as enthusiasm, interest, persuasiveness and nervous energy, body movements such as gestures, body movement mimicry or rhythmic patterns, or proximity of attendees. This information is displayed on mobile phones in an aggregated way. Each group member is represented by a colored rectangle on the screen. A circle is connected to the rectangles via lines. The position of the circle denotes balance of participation, line-thickness reveals speaking times. In their study, the authors compared groups that were supported with feedback to groups without feedback both in co-located and distributed situations. In summary, results indicate that groups are more polite and collaborative with support of the group mirror. Furthermore, introducing the MEETING MEDIATOR reduced differences between co-located and remote situations, especially by distributing the energy of dominant persons among the whole group.

Streng et al. (2009) present an evaluation of different types of visualizations for group mirrors. They compared the more common approach of using an abstract visualization in form of a diagram with the novel approach of metaphoric visualizations (see Figure 2.4, bottom middle). In that system, group members that have the role of reviewers of a specific collaboration script are represented with trees that flourish or fade dependent on the performance of group members. The weather represents the participant with the role of the analyzer, that is, good weather means high quality argumentation. This aspect is the second novelty of this system. Compared to previous work on group mirrors, this system provides feedback on qualitative aspects rather than on quantitative measurable information such as speaking times. The results of the study show that metaphoric feedback is more popular and leads to better self-regulation than the corresponding diagrammatic version.

Brandon et al. (2011) enable group members to interact with the group mirror directly. They developed a system with private displays and with a public display. Group members can draw their own avatars, which are then shown on the displays, surrounded by white circles. These circles grow dependent on the speaking time and connecting lines between the avatars appear representing the turn-taking behavior. The interactive part lies in a functionality that allows participants to change the positioning of their avatar on the screen. Moving it near another person's avatar means consensus with their expressed opinion.

Schiavo et al. (2014) took another step in the direction of how different visual designs can affect the impact of group mirrors. They compared a subtle version with an overt one. The subtle visualization displays on whom was looked the least in a discussion, using colored bubbles (see Figure 2.4, bottom right). These bubbles change their color dependent on which person is currently attended the least and move in his or her direction. The overt version uses text messages that are displayed to the group members to convey this information. Results of their quasi-experiment indicate that subtle directives are more effective than overt ones, however, only if participants understand how they come about. In a second study, Schiavo et al. (2016) added a shared display to the setup. They investigated the influence of social traits in more detail. Results show that participants with a low extroversion score tend to prefer the overt directives while group members with low consciousness trait scores perceive the subtle version as more influential.

A recent approach by Yoshida et al. (2016) uses physical buttons that allow group members of a brainstorming to provide positive feedback. When pressing a button, a sound is played. The main intent of this tool is to decrease negative effects related to evaluation apprehension. In a field study, they could show that the systems increases awareness of positive feedback through sound and the way and frequency the buttons are pressed.

Other recent approaches aim at making the employment of group mirror systems easier by using tablets to mirror speaking times and gaze behavior (see Adachi et al., 2014, 2015). This setup does not require additional software or hardware, since tablets can serve both as sensors and as visualization devices.

In conclusion, previous studies on group mirrors could show that displaying feedback about participation levels can effectively lead to more balanced contributions. Dependent on the specific group mirror visualization, the authors report about different constraints. Some studies have shown an effect only for "over-participators", other studies detected this effect for group members alone who believed that balanced participation was important for the task at hand. While some results indicate that the group mirror was well accepted by the participants, others were perceived as distracting from the conversation. While the accuracy of the feedback seems to be less crucial, feedback solely on the topic could not obtain the same effects as feedback about social aspects. First results on the type of the visualizations speak for subtle visualizations that move away from diagrammatic forms.

Group mirrors for co-located collaboration are especially relevant in the scope of the current thesis. However, a variety of feedback systems supporting lectures or remote collaboration are also important and serve as inspiration and basis for several of the group mirrors that are presented in this thesis. Subsequently, I will review an excerpt of these systems.

2.2.3 Feedback in Lectures

Technical systems used in classrooms or lectures are referred to as classroom communication systems (CCS) (Beatty, 2004). These systems can enhance communication between audience and lecturer in a more unobtrusive way than face-to-face communication. An important part of these tools is the feedback that is provided to both the audience and the lecturer.

A number of systems allow students and lecturers to exchange questions and answers, either as quizzes or in a more open-ended form (see e.g. Dufresne et al., 1996; Harry et al., 2009; Ratto et al., 2003; Reinhardt et al., 2012; Wessels et al., 2007; Yourstone et al., 2008), others permit different forms of note-taking (e.g. Kam et al., 2005; Simon et al., 2008; Wilkerson et al., 2005) or written feedback (e.g. Anderson et al., 2003). An overview of Audience Response Systems (ARS), also called "clickers", has, for instance, been presented by Caldwell (2007), an overview of learning dashboards, categorized in systems supporting face-to-face lectures, supporting group work during face-to-face sessions and online or blended learning scenarios can be found in Verbert et al. (2014). As in the scope of this thesis, the interest lies more on visually provided feedback of collaborative activities, I will outline some of the systems with elaborate visual feedback in the following.

Chen (2003) presents a system for remote lectures using a multiparty videoconferencing system. Below the videos, speech activity is indicated with yellow bars, hand motion with red and body motion with green bars. Aside from that, a summary of the whole conversation is visible (see figure 2.5, top left). Results of their study reveal that teachers cherish especially replay feedback, meaning feedback that is shown following the conversation in contrast to feedback that is shown in real-time. When students were asked if they would feel comfortable if video, audio or the activity information were recorded, most of them spoke in favor of the activity information as this maintains privacy to a higher level than video.

Sturm et al. (2006) developed a system dedicated to provide feedback about the audience to the lecturer. They used a similar color scheme, however in their system, designed for co-located lectures, these colors indicate attention level in form of a pie chart (green = ready, red = busy, grey = other) and interest level in form of a scale (red = low, yellow = medium, green = high) (see figure 2.5, top middle).

Bry et al. (2011) and Pohl et al. (2011) developed the system BACKSTAGE, a digital backchannel used in large lectures and classes. It aims at improving awareness of both the audience and the speaker and at helping students to actively participate in the lecture.

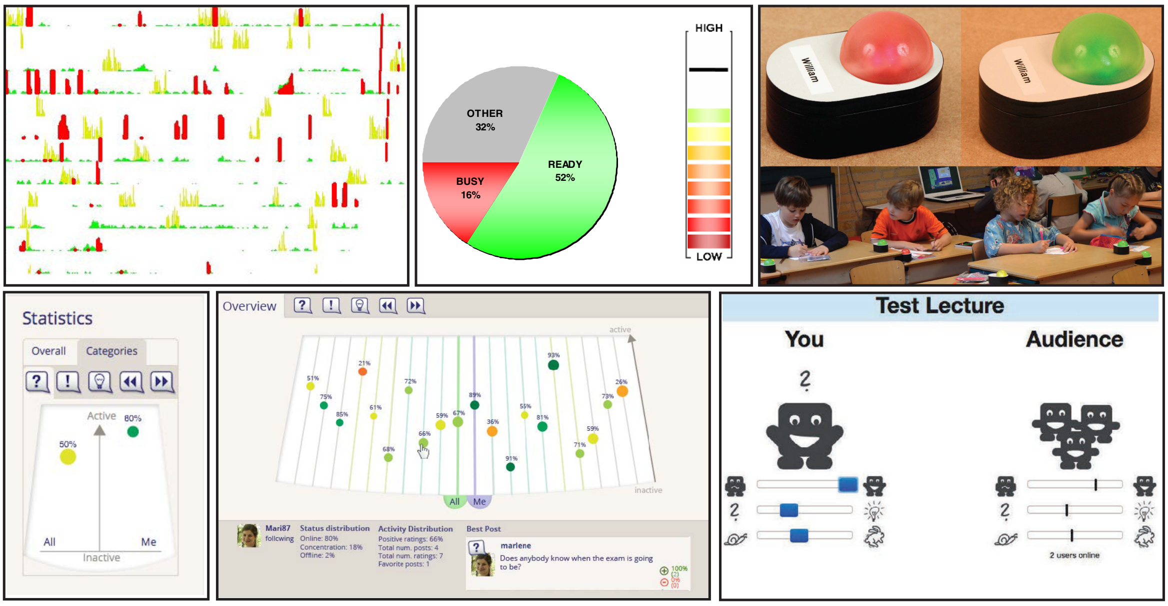


Figure 2.5: Related work on feedback in lectures. Top row from left to right: video conferencing system by Chen (2003), feedback about the audience by a system of Sturm et al. (2006), a feedback system for the classroom by Bakker et al. (2012). Lower row: ACTIVITY AGGREGATOR by Pohl et al. (2012) in the lean version (left) and extended version (middle) and a feedback system by Ebner et al. (2014).

One of the main functionalities is a microblog that enables students to exchange comments and ideas. This can be done publicly or privately as well as anonymously or by using pseudonyms. Students have the possibility to categorize and rate their messages. The best rated messages will then be displayed to the lecturer. The lecturer in turn has the possibility to conduct short quizzes. In a first study, Pohl et al. (2012) could show that with BACKSTAGE, more questions were raised during lectures than without the system. Furthermore, the user study revealed that the system could be improved by integrating presentation slides. This was included in an enhanced version (see e.g. Baumgart et al., 2011; Gehlen–Baum et al., 2012).

Pohl et al. (2012) present a further iteration of the interface. Additionally to the microblog and the slides, there is a space for the virtual representation of the lecturer and awareness components. The interface is designed to support both social as well as workspace awareness (for more details on awareness see Section 2.3). Social awareness is promoted through an indication of the presence of the user (online, offline or busy). The display of certain awareness information is dependent on the mode. Students in busy-mode who want to concentrate on the lecture are only provided with the slides and quizzes. The ACTIVITY AGGREGATOR is a visualization that displays the activity level derived from writing and rating of messages for the student and the audience. The lean version is designed for using it during the lecture (see figure 2.5, bottom left), the extended version as a replay (see figure 2.5, bottom middle). Concepts to support workspace awareness are, for example, the categorization of posts, which reveals strongly debated topics, a notification mechanism that shows all activities related to a student's post, or rating and ranking of the posts. The annotating functionality provides awareness information through the position of the annotation on the slide and through the chosen category. In another study, Gehlen–Baum et al. (2014) compared the use of mobile phones in lectures with and without the support of BACKSTAGE and could show that the system helped students to stay involved with lecture-related activities.

An example of a tool envisaged for the use in classrooms to support communication between teacher and students is the system FIREFLIES by Bakker et al. (2012). It is composed of light objects placed on the students' desks and a teacher-tool with which the teacher controls the light objects by switching between four different colors (see 2.5, top right). Auditory feedback can additionally be provided about the color of the objects (e.g., ocean-sounds representing yellow, owl-sounds representing green). Bakker et al. (2013) evaluated their open-ended interactive system in a study in four different classrooms and observed that it was used in different ways to provide feedback. It was, for example, used to give compliments or to give a turn to a student. In general, the visual feedback was perceived as more successful than the auditory feedback.

Ebner et al. (2014) present a system with which students can provide feedback about satisfaction, understanding and pace autonomously to the lecturer using an interface with sliders. The value of the aggregated feedback is represented through facial expressions and gestures of avatars of the student and of the audience (see figure 2.5, bottom right).

In summary, visual feedback that is provided in lectures to either students or lecturers often uses color coding to transmit information (see, for instance, also a tool presented by Maldonado et al., 2012). Other mechanisms for coding feedback information are using traditional visualization methods such as scales, bar charts or pie charts. One system uses scales in combination with avatars that represent the students. In contrast to group mirrors, these systems are designed for the specific context of the classroom or lectures, in which a clear differentiation between the teacher or lecturer and the students is made. In most cases, different information is presented to the lecturer and to the students.

2.3 Awareness Systems

Another closely related research field are awareness systems, a topic that is of importance for the research areas of CSCW and groupware systems (e.g. Dourish and Bellotti, 1992; Gutwin and Greenberg, 1998; McDaniel and Brinck, 1997; Rodden, 1996). In the following, definitions for the term awareness are given and a short historical overview of research on awareness systems is provided. Afterwards, I will clarify the connection of awareness systems to the topic of this thesis and will give examples of awareness systems that are related to research on group mirrors.

2.3.1 Definition of Awareness Systems

Awareness systems are *"systems intended to help people construct and maintain awareness of each others' activities, context or status, even when the participants are not co-located."* (Markopoulos and Mackay, 2009). In this context, the term awareness needs to be defined. Different definitions exist, for instance, by Endsley (1995) who defines it with the terms *"knowing what is going on"* or by Markopoulos and Mackay (2009), defining it as the *"understanding regarding what others do, where they are, or what they say."*

A variety of different types of awareness have been determined, such as social awareness, describing *"awareness about the social situation of the members"* of a group (Tollmar et al., 1996). Other examples are contextual awareness (Mark et al., 1997), situation awareness (Adams et al., 1995) or workspace awareness (Gutwin and Greenberg, 1996). Situation Awareness (SA) *"refers to the up-to-the-minute cognizance required to operate or maintain a system"* (Adams et al., 1995). Gutwin and Greenberg (2002) see workspace awareness as a specialization of situation awareness and define workspace awareness (WA) as *"the collection of up-to-the minute knowledge a person uses to capture another's interaction with the workspace"* (Gutwin and Greenberg, 1996). A framework of workspace awareness for real-time groupware has been presented by Gutwin and Greenberg (2002).

Kirsch-Pinheiro et al. (2003) present a number of characteristics of awareness systems (for the original version see Kirsch-Pinheiro et al., 2001), organized into six questions that have to be considered interdependently for synchronous and asynchronous environments: **What** information is presented, **when** it its presented, **where** and **how** it is produced and presented, **who** is working and **how much** information should be given. More reviews and discussions of awareness systems can be found, for instance, in the work of Schmidt (2002) or Markopoulos and Mackay (2009). An overview of awareness systems in the context of Computer Supported Collaborative Learning (CSCL) is provided by Janssen and Bodemer (2013). Discussing all existing awareness systems would go beyond the possibilities of the current thesis. Thus, a number of awareness systems are outlined that are interrelated with the group mirrors presented in this thesis in some way, considering that they present feedback about group processes to the collaborators.

To my knowledge, a clear distinction between the terms "group mirror" and "awareness system" does not exist yet. Both terms are used in literature for similar purposes. Jermann and Dillenbourg (2008) coined the term "mirroring systems" or "group mirrors" and also presented such a tool. This tool was later referred to as "awareness system" (Janssen and Bodemer, 2013). The best approach is probably to define group mirrors as a subcategory of awareness systems - though there are several pieces of literature on group mirrors that do not mention the term "awareness" at all. However, group mirrors in general provide information to a group that increases their awareness of these aspects. In that terms, a group mirror is an awareness tool, specifically designed for collaborative situations to make groups aware of ongoing group processes.

2.3.2 A Brief Overview of Awareness Systems

Rittenbruch and McEwan (2009) categorize research on awareness systems historically into three phases. They call the first phase from approximately 1990 to 1994 "early exploration of awareness". In this phase, researchers discerned real-world evidence for the need of awareness in distributed environments. First media spaces were implemented and evaluated such as at the Xerox PARC (Stults, 1986) or successors such as Portholes (Dourish and Bly, 1992), a system which displays images of offices in different locations that are updated every few minutes.

The second phase from 1995 to 1999, "diversification and research prototypes" (Rittenbruch and McEwan, 2009) resulted in a number of new concepts and terminologies regarding awareness. The social context was incorporated and several collaborative environments were developed and evaluated. Examples of systems that are important particularly in the scope of this thesis originating from this phase are ambient awareness systems. There are a number of definitions of ambient awareness systems, for an overview see Pousman and Stasko (2006). Characteristics of the definitions are that the displayed information perceivable in the periphery of the attention is *"important but not critical"* (Pousman and Stasko, 2006), is *"abstract and aesthetic"* (Mankoff et al., 2003) and can move to focus of attention and back to the periphery (Pousman and Stasko, 2006). An early example of such a system is the AMBIENTROOM by Ishii et al. (1998), in which information is encoded in ambient light, sound, water and tangibles.

Referring to Rittenbruch and McEwan (2009), the third phase started around 2000 and is called "extended models and specialization". Work on models and concepts continues, and existing models are refined and extended. Further, new domains such as domestic or medical settings are explored. New technologies and concepts such as tabletop systems and mixed presence collaboration have led to novel developments in awareness research.

2.3.3 Group Feedback in Awareness Systems

In the last two paragraphs, I outlined a quite general introduction to awareness systems, including, for instance, systems that provide awareness information about energy consumption or stock market trends. Awareness information in the context of CSCW and groupware is more related to the topic of this thesis, however this also encompasses systems that, for example, show awareness information about the presence of others in a distributed environment, something that is not necessary in co-located environments as I discuss them in this work. Thus, examples of awareness systems are outlined that focus on making groups aware of group processes during collaboration in the following.

An early example is CHAT CIRCLES (Donath and Viégas, 2002; Viégas and Donath, 1999), a system to enhance presence and activity awareness in an online chat. The tool represents each person by a colored circle that is accompanied with the participant's name and can be moved on the screen to build clusters (see Figure 2.6, top left). Activity is represented by the

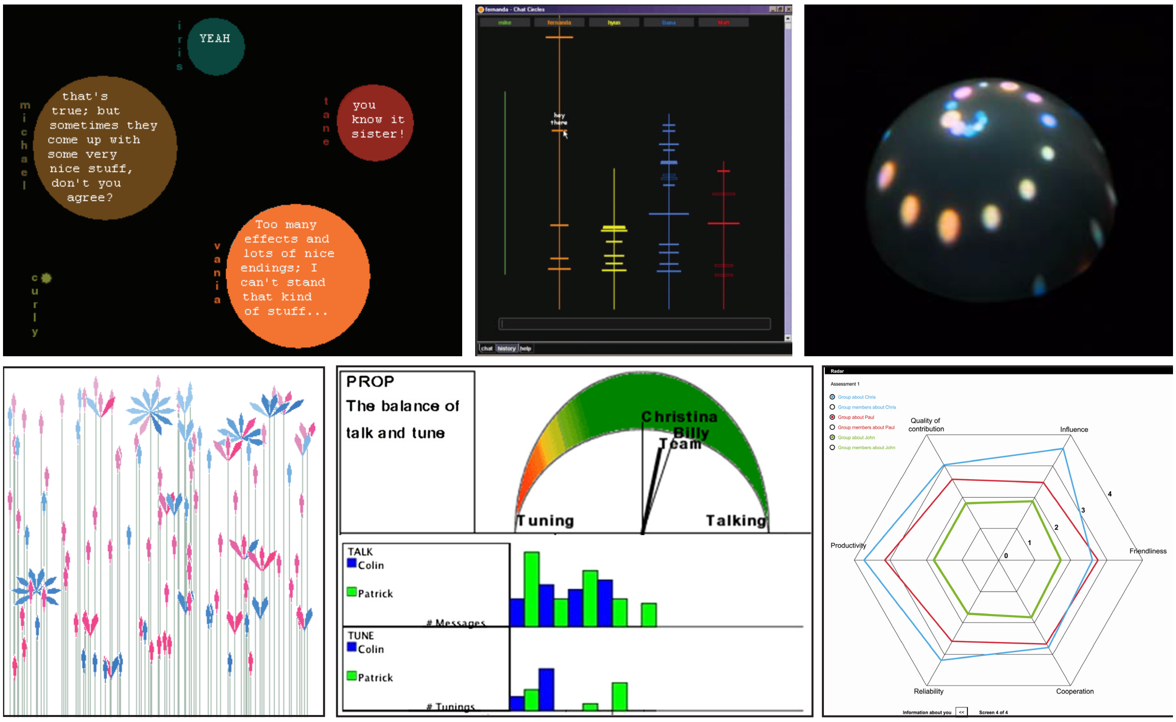


Figure 2.6: Related work on awareness systems. Top row from left to right: CHAT CIRCLES by Donath and Viégas (2002), CONVERSATION LANDSCAPE from the same system displaying the history of the chat, VISIPHONE by Donath et al. (2000). Lower row: a balanced conversation displayed with PEOPLEGARDEN by Xiong and Donath (1999), TALKTUNE PROPORTION metacognitive tool (top) and mirroring tool in form of bar charts (bottom) (Jermann and Dillenbourg (2008)), RADAR by Phielix et al. (2010).

size of the circles. With a posting, the corresponding circle grows. After a while, the posting fades out and the circle shrinks. The brightness of the circles shows how active participants are. Additionally, a visualization of the history called CONVERSATION LANDSCAPE is designed to reveal identity as well as social patterns (see Figure 2.6, top middle). The y-axis represents the time and messages are displayed on this axis as colored bars while the length of the bars represents the length of the messages. For an overview of systems building on CHAT CIRCLES, such as TALKING IN CIRCLES (Rodenstein and Donath, 2000), see Donath and Viégas (2002).

A system designed to assist in audio communication is the VISIPHONE by Donath et al. (2000). It intends to support communication between two people who both have VISIPHONE stations in form of a translucent dome (see Figure 2.6, top right). Information about the audio is displayed as colored dots that spiral outwards on the dome. Different colors are used for the two persons and for silent phases. By that, people can observe if the connection is established and if the audio signal is being transferred. The dome serves as an aesthetic object that shows the rhythm of a conversation.

Xiong and Donath (1999) and Donath (2002) use the metaphor of a garden with blooming flowers to visualize conversation. Group members of an online chat are represented with flowers whose petals represent their postings. Information that is encoded is the time of the posting (petals are arranged by time in clockwise order), the amount of responses (blue petals) and whether a new conversation is started with a post (red petals) (see Figure 2.6, bottom left).

Other systems allow group members to rate contributions of their peers or to indicate understanding. Buder and Bodemer (2007) developed a system tailored to support CSCL discussions. Group members can rate how much they agree to contributions of others and how novel they estimate them. This information is displayed on a two-dimensional graph with agreement and novelty on the axes. Dehler et al. (2011) present a system that mirrors the group members' knowledge. In contrast to Buder and Bodemer (2007), they could show that the tool guided communicative activities.

A system displaying qualitative information is TALKTUNE PROPORTION (Jermann and Dillenbourg, 2008), showing the balance between talking and using a simulator and providing information whether the balance is acceptable (see Figure 2.6, bottom middle). Results indicate that a metacognitive tool (upper image) can effectively influence group processes, and that it is more successful than the compared mirroring tool (lower image).

RADAR and REFLECTOR (Phielix et al., 2010, 2011) are a combination of tools supporting computer supported learning. RADAR is a peer feedback tool that allows group members to rate themselves and others in terms of friendliness, cooperation, reliability, productivity and in terms of the quality of their contributions. The ratings are mirrored to the group in form of a radar diagram (see Figure 2.6, bottom right) so that group members can compare themselves with the group. The REFLECTOR is used to reflect about the own and the group's behavior by answering different reflective questions. In a study it could be shown that with this tool, group satisfaction could be increased.

A number of other awareness systems such as the GROUPMETER (Leshed et al., 2007, 2009) and an approach by Mathur and Karahalios (2009) or awareness tools in the context of computer-supported learning (Janssen and Bodemer, 2013) are also related to the topic of this thesis, but cannot all be described in detail here.

There are several parallels and also mutual influences between the presented awareness systems for distributed group work and the group mirrors that are designed for co-located collaboration. The history visualization of the CONVERSATION CLOCK, for example, resembles the visualization of the history of CHAT CIRCLES. In both systems, colors represent different participants on a temporal axis (in case of the CONVERSATION CLOCK, this axis has a circular shape). One difference is that the length of the bars represents the length of the message in CHAT CIRCLES and the loudness in the CONVERSATION CLOCK. Similarities to the VISIPHONE are, for instance, that the size of the circles represents the volume of the audio and that silent phases are shown using grey dots. The tool TALKTUNE PROPORTION resembles the SECOND MESSENGER, as colored bar charts are used to visualize the amount of speech (in the TALKTUNE PROPORTION, the interactions with the simulator are additionally displayed as bar charts). The group mirror of Streng et al. (2009) uses metaphors such as trees and the weather to represent group processes, Xiong and Donath (1999) use the metaphor of flowers and a garden to reveal conversational patterns.

Systems that support awareness in distributed environments often focus on regaining awareness about information that is getting lost due to the nature of remote collaboration. In face-to-face collaboration, group members are commonly more aware of each other. However, even in face-to-face collaboration there are processes and group dynamics that stay unnoticed by the group. In these cases group mirror systems for co-located collaboration can provide further support. These systems - that have been investigated far less than systems designed for distributed collaboration - are in the focus of this thesis.

3
Application Areas

For the evaluation of technologically mediated feedback on group processes, this thesis focuses on two specific classes of tasks: collaborative creativity and argumentation. These tasks were chosen as they represent two different manifestations of a continuum of tasks reaching from more open-ended to more structured. Though collaborative creativity methods can follow certain structures, the result of this task is undefined and the creative process is difficult to understand and capture. In contrast, argumentation and debates are more structured tasks. The goal is to build proper arguments, for example, with the goal to persuade others of the argument. The next two sections provide an overview of background and related research both about collaborative creativity and argumentation.

3.1 Creativity

Creativity is crucial for the success of both individuals as well as groups and as organizations. Individuals use creativity in different ways in order to solve problems of daily life or to create art. Creativity on a societal level may lead to new inventions, processes, scientific findings or movements in art (Sternberg, 1999). In the 1950s, creativity emerged as a topic of research. Since then, researchers were discordant about how to define *creativity*, some defining the creative process, others the creative product. Approaches were undertaken to define creative individuals, however leading to inconsistent results (see e.g. Barron and Harrington, 1981; Batey and Furnham, 2006).

Taken together, several definitions include the two aspects of "novelty" and "usefulness". Amabile (1983) refers to creativity on a conceptual level and defines that *"a product or response will be judged as creative to the extent that (a) it is both a novel and appropriate, useful, correct, or valuable response to the task at hand and (b) the task is heuristic rather than algorithmic"*, since the outcome cannot be creative when using algorithmic methods which define the process. Sternberg (1999) defines creativity as *"the ability to produce work that is both novel (i.e., original, unexpected) and appropriate (i.e., useful, adaptive concerning task constraints)."*

Summaries of creativity research can, for example, be found in two handbooks, the *Handbook of Creativity* by Sternberg (1999) and the *Creativity Research Handbook* by Runco (1997). A thorough *encyclopedia of creativity* with articles from researchers from a variety of different domains has been published by Runco and Pritzker (1999). Shneiderman (2000) summarizes relevant research in the field of HCI. He proposes a framework consisting of four phases, (1) *collect*, (2) *relate*, (3) *create* and (4) *donate*, including eight activities such as *"searching and browsing digital libraries"* or *"visualizing data and processes"* which Shneiderman identifies as key challenges for HCI research and user interface design.

Other formalizations of the creative process are for instance the combination of divergent and convergent thinking (Farooq et al., 2005; Guilford, 1983). While divergent thinking describes the ability to find several responses or solutions to open-ended questions and tasks, convergent thinking means to find one answer or solution. Csikszentmihalyi (1990) defines five steps of the creative process: *preparation*, which includes becoming immersed in the question or task; *incubation*, which is the process that happens subconsciously during the times a person is not actively engaged in the task; *insight*, which is also refereed to as "aha experience" and in which a person reaches such an insight, *evaluation*, which means that the insight is examined in terms of value and usefulness, and *elaboration*, which comprises to realize the insight. Researchers identified a number of criteria that can facilitate the creative process, such as "clarity of goals" or "reflexivity" (Csikszentmihalyi, 1990), one of them being collaboration (for an overview see e.g., Paulus and Yang, 2000).

The next sections deal with the aspect of collaboration in creativity. First, the two collaborative creativity techniques that have been selected for in-depth analysis in thesis, brainstorming and the Disney Method, are outlined. Then, the aspect of computer support for collaborative creativity is investigated in more detail and examples from related work are provided. Related work about the effects of feedback on creativity is summarized and finally, several approaches are outlined on how to evaluate creativity.

3.1.1 Creativity Techniques

In this thesis, we investigate group mirror systems for two creativity techniques, brainstorming and the Disney Method. Brainstorming (Osborn, 1953) is probably one of the most common techniques and one that has perceived considerable attention in research. The Disney Method (Dilts, 1995) is less known and is based on the concept of different roles to facilitate the flow of creativity.

Brainstorming

Initially invented by Osborn (1953), brainstorming has evolved to a term that is used for a variety of different kinds of creative collaboration. Furthermore, a number of variant forms and supplementations have emerged, for example, *brainwriting* (Geschka, 1978).

However, in its initial form, a process and a number of rules were determined. The process consists of a *storming* phase, in which ideas are generated, and a *norming* phase, in which these ideas are categorized. The technique relies on four principles: (1) *criticism is not permitted*, (2) *unusual ideas are welcome*, (3) *the more ideas the better* and (4) *build on ideas of others* (Osborn, 1953). Two rules are often added: (5) *do not interrupt each other* and (6) *stay on topic* (see e.g., the brainstorming rules proposed by IDEO[1]). The philosophy behind this method and the rule of focusing on quantity rather than quality of ideas is built on the assumption that the large number of ideas will lead to the generation of high-quality ideas as well.

The Disney Method

The Disney Method is based on Walt Disney's way of working and thinking. Dilts (1995) describes this technique as an interplay of three conceptual positions that Walt Disney used. The main idea of the Disney Method is to look at ideas from different perspectives, similarly to the Six Thinking Hats Method (De Bono, 1985). With that, groups can think of novel ideas but also consider feasibility and a more critical view on their ideas. The three roles are:

Dreamer The role of the dreamer is used to produce new ideas and new goals. This role aims at answering the question "what" and provides a vision. It has the bigger picture in mind and is meant to think about ideas in the long term.

Realist The realist concertizes the ideas and fantasies of the dreamer by answering the question "how". This role considers a shorter time frame than the dreamer.

Critic The role of the critic identifies possible problems and addresses constraints both in short and in the long term. By that, it aims at answering the question "why".

Walt Disney designated three different rooms for the three conceptual positions. He would go into these rooms when he wanted to think in one of these three positions. Three different body postures are described that should facilitate the thinking process in the different roles.

3.1.2 Computer Support for Collaborative Creativity

Creativity very often arises through social interaction and collaboration. Csikszentmihalyi (1990) claims that *"an idea or product that deserves the label 'creative' arises from the synergy of many sources and not only from the mind of a single person"*. Summaries of research approaching group creativity have for instance been published by Paulus and Nijstad (2003). Increasingly, collaborative creativity is supported by technology, especially in distributed situations in which collaboration is accomplished over distance. Systems supporting distributed communication are called "Group Support Sytems" (GSS). GSS that are designed to support creative tasks are referred to as "Electronic Brainstorming Systems" (EBS) (Nunamaker et al., 1991).

[1] https://challenges.openideo.com/blog/seven-tips-on-better-brainstorming, last retrieved 17.05.2016

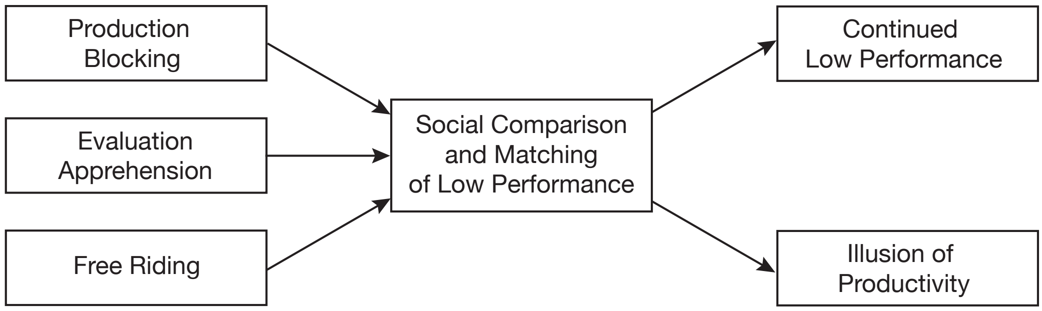


Figure 3.1: Social influence model. A model of productivity losses in brainstorming by Paulus and Dzindolet (1993)

.

These systems bring a number of advantages, but also introduce new problems to collaborative creativity. To discuss problems and challenges, I will firstly leave out technological aspects. Well-known problems of collaborative creativity in general have been described by Diehl and Stroebe (1987). They compared group brainstorming with individual brainstorming and identified the issues of *production blocking*, *free riding* and *evaluation apprehension* in this context. *Production blocking* arises from the fact that only one person of a group can speak at the same time. Other group members might forget their ideas or suppress them as they perceive them as less relevant after a certain time. *Free riding*, the effect that group members “free ride” on the effort of others, can occur when group members feel that the results of their work are monitored on a group level rather than on an individual level. Another reason might be that in bigger groups, the perceived effectiveness of the individual contribution is lowered and group members perceive their input as dispensable. *Evaluation apprehension* names the effect that group members fear the evaluation of other group members and therefore restrain themselves from contributing their ideas. Paulus and Dzindolet (1993) conducted a study in which they identified *social matching* and *comparison with low performers* as one of the main factors for low productivity and the illusion of high productivity (see Figure 3.1). However, collaboration can also ameliorate creativity through synergy effects (Dennis and Valacich, 1993; Lamm and Trommsdorff, 1973) and social facilitation (Zajonc, 1965).

EBS can decrease some of the problems that arise through collaboration in creative tasks. In distributed settings, parallel input prevents *production blocking* and anonymity is a factor that can reduce *evaluation apprehension* (Connolly et al., 1990; Cooper et al., 1998). DeRosa et al. (2007) present a meta-analysis of studies on EBS and confirm a number of these advantages. However, technology also introduces new problems to collaborative creativity. Fischer (2004) identifies a number of barriers to collaborative creativity (in their words “social creativity”) that at the same time introduce opportunities for the design of socio-technical systems. These are aspects concerning distance, meaning not only *spatial* distance, but also *temporal*, *conceptual* and *technological* distance. Especially *spatial* distances, meaning that collaborators are located in different physical locations and *technological* barriers, meaning distances between persons and artifacts, are introduced by EBS.

One large group of EBS are systems that support distributed collaborative creativity (e.g. Gallupe et al., 1991; Sosik et al., 1998; Valacich et al., 1992), bringing along the advantages and disadvantages described before. However, face-to-face brainstorming is not likely to be completely replaced by EBS (Dennis and Reinicke, 2004). As successful group work does not evolve automatically (Cohen, 1994), it is also important to assist groups when working together in a co-located scenario. With the introduction of technology that blends into the environment and everyone's daily life, one can observe a *"paradigm-shift from human-computer interaction to computer-mediated human-to-human interaction"* (Hilliges et al., 2007). This opens the opportunity to design socio-technical systems that combine the advantages of technologically supported communication and collaboration and the advantages of natural face-to-face interaction.

Hilliges et al. (2007) present an overview of the differences between four possibilities of brainstorming support: (1) *PC-based face-to-face*, (2) *electronic remote*, (3) *manual* and (4) *electronic face-to-face*. A main difference between electronic remote and electronic face-to-face collaboration is an increased group awareness with the tradeoff of less anonymity, which has proven to reduce evaluation apprehension (Connolly et al., 1990). In contrast to *PC-based face-to-face* scenarios, *electronic face-to-face* enables natural face-to-face communication, however, at the cost of increased production blocking.

As in this thesis the focus lies on systems supporting co-located collaborative creativity in a subtle and unobtrusive way, the category of *electronic face-to-face* systems is discussed more detailed in the following.

3.1.3 Systems Supporting Co-Located Collaborative Creativity

A common way to support co-located collaborative creativity is to use large displays. This has been realized, for example, with the tools CLEARBOARD (Ishii et al., 1993), POST BRAINSTORM (Guimbretière et al., 2001), IDEAPLAYGROUND (Perteneder et al., 2012) or the TRAINS OF THOUGHT system (Jaco et al., 2014). Other systems combine different types of displays. One of the first systems supporting co-located collaborative creativity and a predecessor of many other collaborative environments combining different input and output devices is i-LAND (Streitz et al., 1999). It is composed of an interactive table and wall and chairs with built-in private displays. Other examples are TEAM STORM (Hailpern et al., 2007) or a system by Hilliges et al. (2007), which I will describe more detailed in this section. Systems combining displays with tangible items use for instance physical Post-it notes on a large wall display (e.g., THE DESIGNER'S OUTPOST by Klemmer et al., 2001) or by combining pen-and-paper methods with an interactive display (e.g., the MEMTABLE by Hunter et al., 2011; or IDEAVIS by Geyer et al., 2012).

Below, I will first describe an example for an interactive creativity support system using a combination of different display types by Hilliges et al. (2007) in more detail. Afterwards, I will review systems that introduce features that resemble group mirror systems as they reflect group processes to the group.

Figure 3.2: Related work on brainstorming support. From left to right: The brainstorming system by Hilliges et al. (2007), FIRESTORM by Clayphan et al. (2011) and TEAMSTORM by Hailpern et al. (2007).

An environment specifically designed to support brainstorming consisting of an interactive table and a wall has been implemented by Hilliges et al. (2007). They developed a system that supports the convergent phase of brainstorming by using an interactive tabletop display and the divergent phase by using a wall display (see Figure 3.2, left). They use a variant of the classical brainstorming called *brainwriting* (Geschka, 1978; VanGundy, 1988) - a technique in which ideas are written on small pieces of paper and placed in the middle of the table before writing down further ideas. With a set of gestures, group members can create, edit and move ideas on virtual Post-it notes, which are represented both on the table and on the wall, enabling externalization of ideas, linking of ideas and building territoriality on the surfaces. Results of a study comparing their system to traditional brainstorming show that the quality and number of ideas was similar in both conditions. Combined with the advantages of storing ideas and processes and an observed tendency of a higher perceived quality of ideas, the authors could demonstrate the opportunities of socio-technological systems.

Hilliges et al. (2007) derive a number of design considerations for socio-technological environments for collaboration and creativity. These are *pseudo-physicality*, implemented in form of virtual Post-it notes in the presented tool, *meta-physicality*, meaning that virtual objects have a distinct and explainable behavior, even if it deviates from the "natural" behavior of such objects, *seamless social transitions*, meaning that transitions between collaborative and individual work are facilitated, and finally, *visibility of social interaction*, which is realized through the visibility of input actions through body language, and also through the constantly updated output representations in the system of Hilliges et al. (2007).

Another example of a tabletop based tool is FIRESTORM (Clayphan et al., 2011). The authors also designed a system with the goal to support fast and concurrent idea generation, with all ideas being visible on a tabletop display (see Figure 3.2, middle). However, in contrast to the tool of Hilliges et al. (2007), they do not promote orientations or territory building, with the idea in mind that group members should be able to move around the table easily and that everyone should feel free to touch any of the displayed ideas. Therefore, ideas are shown in form of a spiral during the divergent phase (in this paper, as often in the context of brainstorming, the divergent phase is called *storming phase* and the convergent phase *norming phase*). Similar to the tool of Hilliges et al. (2007), the system also supports the norming phase. Ideas can be grouped using a "lasso" gesture. One novel feature is attained through visualizing who created which idea. This is at the same time a typical method of group mirrors to increase awareness of group processes. In this case, group members are represented with different colors. The group can effortlessly estimate, who contributed how much. In the norming phase, it is additionally visible, who contributed how much to a certain category. In a study, Clayphan et al. (2011) compared brainstorming with FIRESTORM to brainstorming with a whiteboard. Their results indicate that the color-coded ideas visible for everyone on the tabletop facilitated awareness and reduced free-riding.

Based on the observations with FIRESTORM, the system SCRIPTSTORM (Clayphan et al., 2014) was developed to investigate whether scripted collaboration can enhance the effectiveness of tabletop brainstorming. *"Scripts structure the collaborative learning process by constraining interactions, defining a sequence of activities and specifying individual roles"* (Dillenbourg and Jermann, 2007). In addition to the phases of *storming* and *norming*, a *reflection phase* was introduced. In all phases, the scripting functions can flexibly be turned off or on. In the idea *generation phase*, the color coding that was already implemented in FIRESTORM can be enabled or disabled leading to an identifiable or anonymous mode. In the idea categorization phase, categories can also be color-coded to identify the author. Other scripts can be activated to select leaders or to choose the way of creating categories. The *reflection stage* provides groups the opportunity to review their processes and the outcome. Group members have the possibility to view statistics about number of ideas, categories, links and touches in the *user statistics widget*. Furthermore, groups can recapitulate the scripting options they chose for each stage in the *item stage choices widget*. The *ideas and categories widget* shows the final categorization together with the time spent in each stage. Finally, the *timeline widget* shows any points in the history.

The feature of color-coding ideas dependent on the originator has also been implemented in an earlier system called *TEAM STORM*, focusing on supporting teams of designers (Hailpern et al., 2007). In this system, group members are equipped with private devices from which they can share selected designs on a wall display (see Figure 3.2, right).

These examples show that features of group mirrors such as feedback about the number of ideas or the interaction of group members have already been integrated into some systems designed for collaborative creativity. First results reveal the potential of feedback on group processes in brainstorming scenarios, though other challenges emerged.

3.1.4 Evaluating Creativity

In the beginning of this chapter on creativity, an approach was undertaken to reference definitions that capture the essence of the various definitions of creativity. When trying to measure creativity, these diverse definition illustrate one of the problems. Linked to that is the phenomenon that people's *"thoughts and actions are guided by personal definitions of* *creativity and beliefs about how to foster and evaluate creativity that may be very different from the theories developed by creativity experts"* (Plucker and Runco, 1998, p. 37).

Despite these difficulties, there exist hundreds of creativity tests and measurements. Overviews of these tests have for instance been published by Kaltsounts and Honeywell (1980), Torrance and Goff (1989) and Cropley (2000). Kerr and Gagliardi (2003) makes a differentiation between tests that measure the *creative process* and tests evaluating *creative persons*. A crucial aspect of the *creative process* is the ability of divergent thinking. Thus, divergent thinking tests are among the most common creativity tests. Popular examples are the Guilford battery of creativity tests (Guilford, 1950) or the Torrance Tests of Creative Thinking (TTCT) (Torrance, 1974), which I will describe briefly as an example of such a test.

In the initial version (the test was renormed several times since then), the TTCT is based on four scores that the test should assess: fluency, flexibility, originality and elaboration. Fluency refers to the number of responses, flexibility refers to the number of distinct categories of responses, originality implies statistical infrequency and with elaboration, the level of detail in the responses is meant. The test is composed of two parallel subtests, namely verbal and figural tests, both existent in two forms (A and B). The verbal subtest encompasses for instance asking questions or making guesses about a drawn scene, improving a given product, thinking of unusual uses of an object and listing all consequences of an improbable situation. The figural subtest is composed of three subtasks: composing a drawing, finishing a drawing and composing a drawing from given lines or circles.

Examples of tests for measuring a creative person are for instance the Adjective Checklist (Gough, 1960) which contain a number of adjectives that are positively correlated to creativity, or the NEO Five Factor Personality Inventory (Costa and MacCrae, 1992), which is based on three factors of a personality model: neuroticism, extraversion and openness (NEO is an acronym for these scores).

Especially important in the scope of this thesis is the rating of the *creative product*, in our case the ideas produced using a collaborative creativity technique. For example, the main goal of brainstorming is to increase the quantity and quality of ideas. Dean et al. (2006) provide an overview of the techniques researchers use to evaluate ideas. They reviewed 90 studies and identified three constructs that researchers use as dependent variables for measuring "creativity of ideas": *idea quality*, *idea novelty* and *idea creativity*. *Idea quality* is understood as applicable to the problem at hand (Aiken et al., 1996), as an effective solution (Valacich et al., 1995) and as implementable (Diehl and Stroebe, 1987). *Idea novelty* is defined as an idea that is rare or uncommon (Connolly et al., 1993), both in the eyes of the rater and compared to the space of ideas. Dean et al. (2006) define *idea creativity* as the combination of *idea quality* and *idea novelty*, thus, a creative idea has to apply to the problem, be effective, implementable and novel. However, several of the reviewed studies define these terms differently.

Dean et al. (2006) discovered four constructs that are commonly used to evaluate ideas: *novelty*, *workability*, *relevance* and *specificity*. Relying on MacCrimmon and Wagner (1994), they define the terms as follows: *"An idea is most novel if nobody has expressed it before"*, it is *"workable if it does not violate known constraints or if it can be easily implemented"*, it *"is relevant if it satisfies the goals set by the problem solver"* and it is *"thorough if it is worked out in detail"*. (Dean et al., 2006 use the term *specificity* instead of *thoroughness*). Dean et al. (2006) provide an overview of studies measuring these factors. From the 90 reviewed studies, 20% counted ideas (e.g., count every entry made in an EBS), 23% used a single holistic measurement (meaning that only a single measure of quality or a single measure of creativity was used), and the remaining papers used a multidimensional measure.

One of the main challenges of all these tests is the scoring of the responses. As mentioned above, the TTCT initially scored the responses using the principles of fluency, flexibility, originality and elaboration, which is traditionally done by raters scoring all answers. Recently, other approaches have been proposed. Silvia et al. (2008) suggest a subjective scoring technique in which participants make a pre-selection of their most creative responses. Plucker et al. (2011) present a comparison of scoring methods and conclude that percentage scoring methods might be most appropriate, a scoring method in which originality scores are divided by fluidity scores. Furthermore, artificial intelligence and computational creativity are recent topics of research. It stands to reason that automated assessment of creativity is also an emerging topic in research, albeit still in its infancy. Maher (2010) and Fuge et al. (2013) present first approaches towards metrics for automatically measuring creativity.

3.1.5 The Effects of Feedback on Creativity

Zhou (1998) conducted one of the first studies specifically dedicated to investigate the effects of feedback on creativity (however, not on collaborative but on individual creativity). She investigated feedback valence (positive or negative), feedback style (informational or controlling) (Pittman et al., 1980), task autonomy (high or low control over how to carry out a task) (Hackman and Oldham, 1980) and the combination of these factors. Feedback was provided in a written form on a feedback sheet. Creativity was measured using a simplified form of a technique by Amabile (1982), using three judges that rated ideas on an 11-point scale. Results showed that positive feedback led to greater creativity as well as informational feedback. Most interesting were the interactions that emerged: with positive feedback in an informational style and with high autonomy, individuals were most creative. The results of the study additionally show the superiority of positive feedback compared to no feedback on creativity.

However, additional challenges exist in the context of collaborative creativity compared to individual creativity. As described in Section 3.1.2, several studies could show that nominal groups outperform real groups in brainstorming sessions (Diehl and Stroebe, 1987; Mullen et al., 1991). Paulus and Dzindolet (1993) conducted several studies to understand the reasons behind that productivity loss better. In one of their studies (Paulus and Dzindolet, 1993, 5th study), they undertook a first approach of investigating the influence of feedback on brainstorming. They compared how nominal groups and interactive groups performed in two different conditions. Those were an “information condition”, in which participants were provided with a standard, meaning that they were told how many ideas other groups/other subjects produced typically and a condition in which participants were not told about this standard. Results showed that the productivity gap could be eliminated when interactive groups were provided with a standard. As this standard is some kind of normative feedback, these results provided first evidence of the potential of feedback in collaborative creativity.

An environment in which creativity often occurs in relation to other people or in collaboration is at work. A considerable amount of research exists that investigates feedback on creativity in work environments. Zhou (2003) and George and Zhou (2007) for instance could show that supervisor developmental feedback could increase creativity of employees. With developmental feedback, the authors refer to valuable feedback that employees can use to learn and improve. Specifically, results showed that creativity increased for less creative personalities in companion of creative co-workers, leading supervisors to provide more developmental feedback. These results indicate that the group composition plays a role in collaborative creativity, as creative individuals can affect the situation. A study by De Stobbeleir et al. (2011) indicates that not only the feedback that is passively received but also feedback that is actively seeked can facilitate creativity. The authors suggest to provide employees with an environment in which they can easily inquire feedback.

I will now turn from these general insights on feedback in creativity to the influence of feedback on creativity in computer-mediated settings. Roy et al. (1996) investigated the effects of feedback in an electronic brainstorming scenario, particularly focusing on the factors of *social matching* and *social loafing*. In their study, groups of 5 to 6 participants generated ideas using an EBS. Each group member was seated at an individual work station in the same room. In the study, three conditions were compared: nominal groups with a public screen displaying the ideas of the other group members (1) during the brainstorming (2) after the brainstorming and (3) without a public screen. It was not possible to see, who wrote which idea. The results of this study reveal that groups who had a public screen (either continuously or afterwards) outperformed the groups without feedback in terms of number of unique ideas. The authors could furthermore show that the continuous feedback induces social matching and thereby equalized performance. These results provided first evidence of the positive influence of feedback on electrical brainstorming.

Besides that, the question remains how computer-mediated feedback works in *interactive* (i.e. co-located) groups engaged in a creative task. To my knowledge, two studies investigated this aspect in more detail. One is a study with the MEETING MEDIATOR (Kim et al., 2008), the other a study by Schiavo et al. (2014). Both systems are described in detail in Section 2.2.2. The study with the MEETING MEDIATOR used brainstorming as a use case, but the displayed information (e.g., speaking time, social signals, body movements) is not necessarily specific to brainstorming. However, in the evaluation, a metric for measuring the success of the brainstorming was used (number of ideas). The number of ideas did not increase due to the feedback of the MEETING MEDIATOR. Furthermore, the results obtained during the brainstorming phase were compared with a problem solving phase. For instance, having a dominant person in a group affected the brainstorming negatively, which was not the case when dominant group members were present during a problem solving task. The authors could observe that this effect emerged because non-dominant group members tended to state less ideas when a dominant person was present compared to groups with only non-dominant participants.

The second approach dealing with technologically-mediated feedback for creative tasks in co-located groups was performed by Schiavo et al. (2014, 2016). One of the main goals of their system was to balance participation during a brainstorming session. In their first study with only private displays, Schiavo et al. (2014) evaluated the differences between subtle directives without explanation, subtle directives with explanation and overt directives. A summary of the results, about the influence of different social traits on the attitude towards the system, for example, can be found in Section 2.2.2. However, results did not reveal significant differences regarding the amount of ideas of the brainstorming. In the second study (Schiavo et al., 2016), a public display was added to the setup. Again, no significant differences were found between the three conditions regarding the number of ideas.

3.2 Argumentation

Argumentation is an important skill. It may be that children want to persuade each other to share their toys, that an employee wants to convince his boss to pursue that specific project he or she is working on, or that a politician wants to propose a novel strategy. Argumentation, both in oral and in written form, is part of everyone's daily life. The attempts of understanding and improving argumentation by defining and evaluating arguments go as far back as to ancient Chinese or Greek culture (Lu, 1998). Aristotle differentiated between certain kinds of argumentation. In the last decades, a growing interest in a variety of domains has evolved such as *"philosophy, logic, linguistics, discourse analysis, rhetoric, speech communication, education, psychology, sociology, political science, law, and many other disciplines [...] and some multidisciplinary and interdisciplinary approaches to argumentation"* (Van Eemeren et al., 1996).

Van Eemeren et al. (1996) define argumentation as *"a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a rational judge."* I will briefly go through the most important terms of this definition and explain them by referring to the explanations of Van Eemeren et al. (1996). Argumentation is always a *verbal* activity, using words and sentences. However, this can of course be accompanied (but not replaced) with non-verbal cues such as gestures or facial expressions. Second, argumentation is *social*, naturally when several people are involved in a discourse but also when a single person weighs up arguments for him- or herself, as this involves contemplating reactions of people. Furthermore, argumentation is an activity of *reason*, meaning that a rational account can be provided for the statements someone brings forward. *Standpoints* can be of different nature, for example, absolute ("It is certain..."), more restrained ("It is likely...") or containing values ("It is good..."). In all cases, standpoints are controversial. *Justifying* a standpoint means to find an acceptable proposition for that standpoint by bringing "pro-arguments", while *refuting* means to show that a proposition is unacceptable and the contrary holds true by bringing up "contra-arguments". Finally, *increasing (or decreasing) the acceptability for the listener or reader* means to persuade the audience with arguments. Thereby, the audience needs to fulfill the role of a *rational judge*.

Especially important in argumentation theory is *"the production, analysis and evaluation of argumentation"* (Andriessen, 2006). One can differentiate between two approaches, the studies of arguments as a more grammatical concept and as a concept relying on the social nature of argumentation (Andriessen, 2006). The probably most well-known approach of defining the "grammar" of an argument and determining the structure of valid and sound arguments is the model by Toulmin (1958). He distinguishes five components of arguments. The *claim* is an expression of position (a standpoint or conclusion), the *datum* is information, facts and opinions that support the claim, a *warrant* justifies the datum as support for the claim (e.g., through definitions, theoretical laws), the *backing* can provide additional evidence in form of specific information supporting the warrant (e.g., statistics, expert opinions), the *qualifier* states the degree of certainty or uncertainty to the claim (e.g., with keywords such as "probably" or "perhaps"), and the *rebuttal* are facts or opinions that question or weaken the claim such as exceptions to the claim.

A simplified model has been published that summarizes the components *datum*, *warrant* and *backing* under the term *ground* and *qualifier* and *rebuttal* under the term *qualifications*, resulting in a tripartite model consisting of *claim*, *ground* and *qualification* (Stegmann et al., 2012; Weinberger et al., 2007).

The second approach incorporates the social component of argumentation. The theory of formal dialectics (Barth and Krabbe, 1982) sees argumentation as a dialogue; the pragma-dialectical perspective (Van Eemeren and Grootendorst, 1992) emphasizes the interaction between proponent and opponent. Another example is the work of Leitão (2000), who strengthens the importance of argument sequences, consisting of *argument*, *counterargument* and *reply*. The *argument "consists of a position which is either followed or anticipated by a justification"* (Leitão, 2000). The *counterargument* then claims something that questions the speaker's position, potentially undermining the position that then needs a *reply*, *"which is designed precisely to capture the arguer's immediate or remote reactions to a counterargument"* (Leitão, 2000).

After this definition and general introduction to argumentation, the next section will describe two argumentation types that are of importance for this thesis: debates and collaborative argumentation. I will then review existing solutions of computer support for argumentation and summarize characteristics of these tools. Systems designed for co-located argumentation are reviewed in more detail. Finally, it is discussed how argumentation can be evaluated and light is shed on the role of feedback in argumentation.

3.2.1 Argumentation Types

In general, there are two types of argumentation: debates and collaborative argumentation (Andriessen, 2006). Both are topic of this thesis, therefore both are described in more detail in this section. For debates, the type of the *British Parliament Style* debate is outlined as an example, because one of the prototypes discussed in this thesis was implemented to support this technique.

Debates

The definition of debates is a slightly different one than the one of argumentation. Freeley and Steinberg (2013) define debates as *"the process of inquiry and advocacy; the seeking of a reasoned judgment on a proposition"* while argumentation is defined as *"reason giving in communicative situations by people whose purpose is the justification of acts, beliefs, attitudes and values."*

A specific debating style that is often used in debate clubs is the *British Parliamentary Style* debate[2]. This type is based on practices of the British parliament system. It consists of four fractions with each two members. Two fractions speak in favor of a certain topic (the "motion"), two against. The fractions speaking in favor of the motion are the opening government and the closing government, the two speaking against, are the opening and closing opposition. The order in which speakers raise their arguments is determined. The first speaker of the opening government starts by defining the motion and by stating arguments, followed by the first speaker of the opening opposition who can react to the arguments by stating counterarguments, but also by producing own arguments. Then, the other speakers follow likewise: the second speaker of the opening government, the second speaker of the opening opposition, the first speaker of the closing government and then the first speaker of the closing opposition. Roughly speaking, these all rebut arguments of the opposing fraction and raise own arguments - with each having a slightly different role. Lastly, the second speaker of the closing government and then the second speaker of the closing opposition have the role - after having the chance to state even more novel arguments - to summarize the debate.

In a debate club, the speakers typically have 15 minutes before the debate to prepare, discuss within the teams of two and take notes that can also be used during the debate. In the course of the debate, each speaker has seven minutes to bring arguments, while the first and the last minute are 'protected', meaning that no questions are allowed during that time. In the remaining five minutes, other speakers and the audience are allowed to ask questions ("Points of Information" (PoI)). For that, members of the audience stand up and raise their hand. The speaker is free to accept and to answer the question or not.

One or several jury members are present during a debate. After the debate, these judges select the winning fraction and justify their decision. Typically, they use evaluation sheets

[2] http://idebate.org/about/debate/formats, retrieved 21.03.2016

tailored for assessing debates. In debate clubs, jury members also use their assessment to provide feedback to the debaters with the aim that these can learn and improve their debating skills.

Collaborative argumentation

As opposed to debates, collaborative argumentation (Andriessen, 2006) sets the focus on agreeing on arguments rather than winning an argument. Andriessen (2006) gives the example of science, where two scientists with different opinions still follow the same goals of resolving the issue and finding an agreement. Students who are a *arguing to learn* serve as another example. The effectiveness of the concept of *arguing to learn* has been proven in several studies. Andriessen (2006) summarizes four reasons. First, elaboration, reasoning and reflection, which are all involved in argumentation, are processes that facilitate a deeper conceptual learning (Bransford et al., 1999). Second, the concept of *learning to argue* is implicitly given. Students do not only learn about the topic at hand but they also learn how to argue (Kuhn, 2001). Third, being collaboratively engaged in argumentation can support general social and collaborative skills such as social awareness (Vygotsky, 1980). Forth, groups often share a "common tradition of argumentation" (Andriessen, 2006) (e.g., in specific domains in science) which requires a specific form of argumentation that people learn through collaborative argumentation (Billig, 1996).

3.2.2 Computer Support for Argumentation

As mentioned above, argumentation is an important skill. But since many people are poor arguers (Tannen, 1998), learning to argue is to be regarded as a favorable goal. However, due to teachers' time and availability constraints, teachings on arguing often fall short. Since the last 20 to 25 years, software tools have been built to fill this gap, for a number of different domains including law (see e.g. Aleven and Ashley, 1997), science (see e.g. Linn et al., 1998), conversational argumentation (see e.g. McAlister et al., 2004) or, as would seem natural, the field of computer supported collaborative learning (CSCL) (see e.g. Andriessen, 2006; Baker, 2003; Scheuer et al., 2010; Stegmann et al., 2007).

Scheuer et al. (2010) published a review on computer support for argumentation. They categorize software tools into tools designed for single users, for small groups and for communities. I will exemplarily describe one tool for each of these categories. However, one needs to be aware that tools might be used differently than they are intended to. Tools for single users might be used by several people sharing a single computer and individuals can use tools designed for collaboration.

LARGO (Pinkwart et al., 2006, 2007) is a tool designed for single users. It is described as an intelligent tutoring system that aims at helping law students to develop legal argumentation skills. Students have to transcribe given arguments from the US Supreme Court into graphical representations (see Figure 3.3, left). The argumentation model this system is built on is a domain-specific model aligned to the requirements of legal argumentation. Feedback

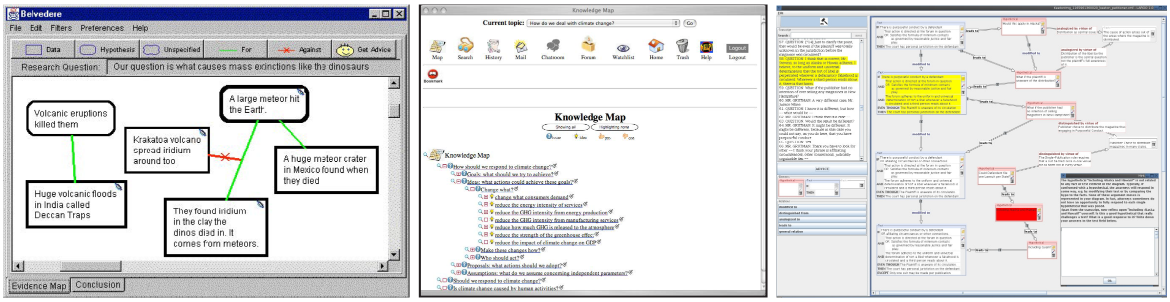


Figure 3.3: Related work on support for remote argumentation. From left to right: LARGO (Pinkwart et al., 2006), BELVEDERE (Suthers et al., 2001 and COLLABORATORIUM (Klein and Iandoli, 2008).

about certain weaknesses (structural weakness, context weakness and content weakness) in arguments is provided in form of self-explanation prompts. This type of feedback is especially helpful in ill-structured domains (such as law) and is aimed at encouraging students to explain their solution.

BELVEDERE (Suthers et al., 2001) is one of the most well-known argumentation support tools and is designed for small groups. Several evolving versions of this tool exist, moving from a focus on scientific reasoning to more evidential argumentation (Scheuer et al., 2010). BELVEDERE is a graph-based system. Nodes represent the statements (classified into “data”, “hypothesis” or “unspecified”) and links can be added to display relations (“for” or “against”). An example is displayed in Figure 3.3, middle. BELVEDERE offers on-demand feedback for students. Similar to LARGO, it shows feedback in form of suggestions and questions, and highlights problematic parts in the graphical representation.

An example for a system designed for larger groups is the COLLABORATORIUM (Klein and Iandoli, 2008). The authors’ main goal was to build a system that combines the advantages of open source/peer production tools with argumentation systems. The resulting system COLLABORATORIUM is web-based and allows people to create argument maps with the aim to support “collaborative deliberation” (Walton and Krabbe, 1995) (see Figure 3.3, right). In this system, moderators evaluate the entries in terms of correctness and validity.

I will now review some of the specific characteristics of these systems that are important in the scope of this theses. Doing that, I will follow the review of Scheuer et al. (2010), who categorized existing argumentation systems. They identified five different **graphical representations** of arguments in the literature: (1) *linear*, (2) *threaded*, (3) *graph-based*, (4) *container* and (5) *matrix*. An example for *linear* is a chat, which is a kind of computer-mediated communication (CMC) in textual form. However, the problem of sequential incoherence exists, meaning that contributions such as questions and answers cannot be matched to the statement they refer to. That is the reason why dedicated argumentation support tools rarely or never use this form. *Threaded* argumentation in contrast solves this problem by enabling message-reply sequences, an option that, for instance, has been made use of in the system HERMES (Karacapilidis and Papadias, 2001). The most common representation form are *graph-based* representations which use nodes and links to display argument components and their relations (e.g., BELVEDERE). Other types are the *container* representation in which elements that are interrelated (e.g., claim and evidence) are contained inside of a frame, and the *matrix* representation that indicates relations inside of the cells. To our knowledge this representation is only used in BELVEDERE, additionally to the graph.

Moreover, Scheuer et al. (2010) discuss the **amount of autonomy** that the systems provides to users when creating arguments. This is especially relevant in the context of learning because a different amount of autonomy might relate to different learning goals. Scheuer et al. (2010) identify five different levels: (1) *free-form arguments*, meaning that users can freely choose the content of their argument components (to a predefined topic), (2) *argumentation based on background materials*, implying that users are provided with background materials, (3) *arguments rephrased from a transcript*, where users are provided with a transcript that they should convey into a more structured form (such as in LARGO), (4) *arguments extracted from a transcript*, meaning that users are able to reuse the wording from the transcript and copy and paste it into a new structure and (5) *system-provided knowledge units*, where the components already exist but the user has the task to relate them to each other.

Finally, it is important to discuss the topic of **ontologies** in the context of argumentation systems, as all systems are based on a specific ontology that may differ from the ones of other systems. *"An ontology is an explicit specification of a conceptualization"* (Gruber, 1993). Traditionally used more in the context of artificial intelligence and knowledge representation, ontologies also take effect in argumentation. In this context, ontologies *"describe the components of arguments, together with relations between components and modifiers of the components and relations"* (Scheuer et al., 2010). In the context of argumentation system, ontologies are used to make users aware of the available conceptual components (Suthers, 2003). Schwarz and Glassner (2007) differentiate between *informal ontologies* and *educated ontologies*. While *informal ontologies* are based on argumentation found in natural conversation, *educated ontologies* contain definitions and rules and result from a more reflected process. Another differentiation in the scope of ontologies is that between rather *general* (e.g., "claim" and "evidence") and *domain-specific ontologies* (e.g., "hypotheticals", a category used in law).

Schwarz and Glassner (2007) state that such ontologies that result from an evolutionary and reflected process can be classified as educated ontologies. They are learned in schools and universities in the form of definitions and rules. This contrasts with informal ontologies, which are based on reasoning that typically occurs in natural conversations. While educated ontologies seem especially appropriate for argument modeling, their informal counterpart may be more suited to support structured - and typically less formal - communication. One variation are sentence opener interfaces, which do not explicitly expose categories but which scaffold new contributions through predefined sentence-starting phrases. Typically, these interfaces are based on an underlying model of desired communication acts and processes, for instance, dialogue games (McAlister et al. 2004). One general problem that communication ontologies and sentence openers strive to address is to help students to stay on topic by limiting user options.

Several of the systems that Scheuer et al. (2010) review are of collaborative nature. All of them are designed for the use on single computers, implying that each collaborator is working from a single computer. Yet in reality, argumentation often occurs in face-to-face situations. I am especially interested in exploring the potential of technological support of such situations. Hereafter, I will discuss first approaches of supporting face-to-face argumentation through technology.

3.2.3 Systems Supporting Co-Located Argumentation

Compared to the huge amount of systems designed for supporting distributed collaborative argumentation, there is only little research on prototypes designated to co-located collaborative argumentation and debate. To our knowledge, there are only two versions of one system, developed at our university, that are specifically dedicated to support face-to-face argumentation and debate (Streng, 2010; Tausch, 2011).

An example are the two versions of one system, the ARGUE TABLE and the ARGUE WALL (Streng, 2010, see Figure 3.4). These systems are specifically dedicated to support argumentative knowledge construction (Andriessen, 2006; Koschmann, 2003). The ARGUE TABLE was the first version of this system. Collaborators could create arguments, constructed of the three parts "claim", "grounds" and "qualification" (Toulmin, 1958) on an interactive tabletop. The arguments could be labeled with a keyword and minimized (only the keyword is visible) and maximized (keyword plus argument components are visible). The three argument parts are made clear by using "sentence openers". The field for the input of a claim is labeled with the sentence opener "I think that...", the field dedicated for the ground with "because..." and the field for the qualification with "unless...". On top of the keywords, the

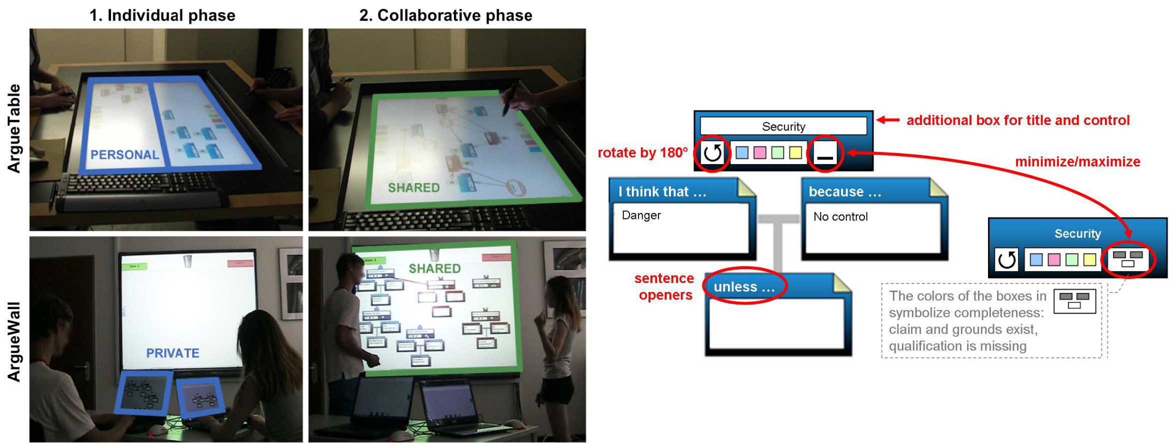


Figure 3.4: Related work on support for co-located argumentation. Left: ARGUE TABLE and ARGUE WALL in the individual and in the collaborative phase. Right: Arguments consist of claim, explanation and example, indicated with "sentence openers". Arguments can be rotated, minimized and maximized (Streng, 2010).

metaphor of puzzle pieces is used to indicate that arguments can be linked to each other, supporting the concept of argument sequences by Leitão (2000). In a study, the authors noticed that in the first phase, in which participants should build arguments, most interaction took part whereas in the second part, in which participants should connect arguments to each other, interaction decreased.

Therefore, a second design was implemented, the ARGUE WALL. This system uses a setup with personal computers and a shared wall display. The personal computers are used for the first phase in which arguments are built, while the second phase in which arguments are linked to each other is conducted on the shared wall display. Streng (2010) compared the two scenarios in a study. Participants should debate about two topics, "pro/con tuition fees" and "pro/con speed limits on highways". Results of their study show that the display setting and the clearer separation of the individual and the collaborative phase with the ARGUE WALL had an influence on argumentation. Working on a table in the individual phase increased awareness between collaborators, but it also led to the feeling of being observed and being disturbed, which was not the case when working on private displays in the ARGUE WALL setup. The collaborative phase was better supported in the ARGUE WALL condition. Participants created more links between arguments. Results indicate that reasons for that were the clear separation between individual and collaborative phase and the better overview on the wall display.

Another system is also related, as the authors investigated a system designed for collaborative argumentation on single computers, but merely in the context of co-located classrooms (Asterhan and Eisenmann, 2011). They describe the differences between more active and more silent students when using an argumentation tool (DIGALO) in a co-located scenario. They could show that participation is more balanced with the tool. Silent students prefer working with this tool while more active students do not have a clear preference.

3.2.4 Evaluating Arguments

In the first place, the produced content of users of argumentation tools has to be analyzed to gather information to provide feedback on. Scheuer et al. (2010) differentiate between argument analysis and discussion analysis. In the following, both approaches and the different possibilities on how to perform these forms of analysis are briefly outlined, following the classification of Scheuer et al. (2010).

Argument analysis focuses on analyzing the structure of arguments. This can be done in different ways, for example, by analyzing *domain-specific patterns*. With this, Scheuer et al. (2010) describe systems searching for specific syntactical patterns that, for instance, violate the argumentation model (e.g., circular arguments). Another possibility is a *problem specific* approach. That technique compares the content produced by the user with content constructed by experts. *Simulated reasoning* leaves out the content of arguments and focuses instead on the network structure. This is analyzed using formal-logical models of validity. For example, in the field of law, this analysis might result in a proof standard such as "beyond reasonable doubt". *Content quality assessment* is currently hard to perform automatically (in contrast to the other approaches described before). Thus, including peers to provide feedback on argument content is a promising strategy, because students do not only think about their own argumentation but reflect also on their collaborators' work. Finally, for gathering data to provide feedback on, it makes sense to be aware of different *phases* of the process (e.g., construction of argument components, linking argument components, error correction etc.).

The second form of analysis in the context of argumentation is a **discussion analysis** (Scheuer et al., 2010), analyzing how collaborators interact with each other. The *discussion process* may be analyzed using classifiers. This means that certain constructs can be detected, both on the "micro-level of argumentation" (i.e., claims, warrants etc.) and on the "macro-level of argumentation" (i.e., arguments and counterarguments etc.). This is carried out using "sentence openers" (e.g., "I think that...", "because...", "unless...") from which students have the opportunity to choose. A second option is to use machine learning techniques that automatically derive classifiers for content that students enter manually. For further analysis, it might make sense to determine the *discussion topic*. This may be achieved through methods known from other technical domains, e.g., knowledge engineering, machine learning or information retrieval. Another helpful analysis is to identify *discussion problems*. Emerging problems are detected, e.g., *"problematic patterns in discussions, failed attempts to share knowledge, and lack of responsiveness"* (Scheuer et al., 2010). With *student and group models*, data on the student's or groups' behavior are collected, aggregated and translated into models. These models allow to show statistics on group behavior, for example, on conversation speed or on agreement inside of a group. Finally, Scheuer et al. (2010) mention the different *discussion phases* that can be analyzed (e.g., in a structured debate, phases in which arguments and counterarguments are produced are strictly regulated).

3.2.5 The Effects of Feedback on Argumentation

The data collected through the forms of analysis described above allow to provide feedback to students and groups. Scheuer et al. (2010) identifies three important aspects of feedback in argumentation. With **timing and control** he refers to the circumstance when the feedback is provided (*on demand*, *immediately* or *summative*) and who provides the feedback (*system*, *peers*, *moderator*). Second, the **mode and content** of feedback plays a role. In argumentation systems, feedback is either displayed in a *textual* form, by *highlighting* certain things, especially in graph-based representations or in form of *meters*. These are mirroring or metacognitive tools (described in detail in Chapter 2). Third, the aspects of **feedback selection and priority** has to be considered, since it is essential to deliberate the amount of feedback carefully, as too much feedback might overwhelm students.

To our knowledge, the aspect of feedback has not been systematically evaluated in many studies. Scheuer et al. (2010) name two studies that investigate adaptive support, one with LARGO (Pinkwart et al., 2007) and one with a system called CONVINCEME (Schank, 1995).

Pinkwart et al. (2007) conducted a study with their system LARGO and compared it with groups using a notepad. As stated above, LARGO is a system that integrates feedback in form of self-explanation prompts. Their results show that for low-aptitude students, the tool provided a number of advantages over the note-taking technique. The authors suggest that one important factor for the improvement of these students is feedback. The authors could observe that the feedback functionality was increasingly used towards the end of the test, indicating that students considered the feedback as helpful. In a second study (Pinkwart et al., 2008), these results could not be replicated. The advice function was used less often and this also decreased over time. One explanation might be that students in the first study (paid volunteers) were more motivated than participants of the second study (mandatory task for students of a class without any additional benefit). However, it is not clear which of the results, if any, can be attributed to the feedback. For that, a study comparing a condition with and a condition without feedback would have been necessary.

CONVINCEME (Schank, 1995) is an argumentation support tool with which students can create arguments and provide ratings on how strong they perceive the individual parts of an argument. Feedback is given based on a model called ECHO. ECHO is a computational implementation of the TEC model (Theory of Explanatory Coherence) (see e.g. Thagard, 1989). Students are enabled to compare their ratings with that computed by the model. In a study, Schank (1995) compared the tool with a traditional pen-and-paper method. Results show that the tool could effectively support students in structuring and also revising arguments. As with the study of Pinkwart et al. (2007), the authors did not isolate the factor of feedback, thus, it is not possible to estimate which part the feedback contributes to the results.

In the last two sections, related research on the two use cases for group mirrors - collaborative creativity and argumentation - have been discussed. In both areas, first approaches of including technologically mediated feedback have been made. These first prototypes and studies reveal the potential of such applications. In this thesis, the influence of several aspects of technologically mediated feedback are investigated in detail. In the next chapters, a design space is presented, followed by the two main chapters presenting prototypes and studies in the context of collaborative creativity and in the context of argumentation.

II

A DESIGN SPACE FOR GROUP MIRRORS

4 Design Space

Research on group mirror systems is a relatively novel field, and therefore lacks a shared understanding of the overall design space. Jermann et al. (2001) and Streng et al. (2009) suggest two different ways of categorizing feedback systems for groups. Jermann et al. (2001) propose a classification framework that distinguishes between *mirroring systems*, *metacognitive tools* and *guidance systems*. Streng et al. (2009) define a design space consisting of the *type of information*, the *type of visualization* and the *placement*. However, these classifications only cover a certain perspective on feedback systems. The following classification provides a more exhaustive view, defining the most important characteristics of group mirrors.

In the next sections, I will outline the existing classifications by Jermann et al. (2001) and Streng et al. (2009) in more detail. Besides that, I will discuss variables of the CSCW design space in regard to their influence on group mirrors. Afterwards, I will define a design space that describes the possible characteristics of group mirrors. In the last section, I will explain and motivate the design choices that have been made throughout this thesis, for example, the focus on co-located scenarios instead of distributed environments.

4.1 Previous Classifications

Jermann et al. (2001) provide a classification of CSCL systems. Their classification focuses on online tools for distributed collaboration in the context of learning scenarios. They distinguish between *mirroring systems*, *metacognitive tools* and *coaching systems*. *Mirroring systems* reflect basic actions to the collaborators or teachers. An example of a *mirroring system* is a visualization of the actions students undertake in a CSCL system, for instance, a visualization of the amount of contributions in a chat. *Metacognitive tools* are systems that compare the current state of interaction to the desired state of interaction. This can be accomplished by visualizing the current state next to the desired state, for example, by showing the current number of contributions in a chat next to the desired amount of contributions. This information is then either visualized and displayed to the participants to promote self-regulation, or it is collected to be analyzed afterwards by a researcher or coaching agent.

Coaching systems use a model of interaction to interpret indicators and offer guidance. In case of the chat example this could be an advice to the collaborators on how to engage more or less in the discussion.

The classification of Streng et al. (2009) applies to group mirrors for co-located collaboration. While the definitions of Jermann et al. (2001) are located on a more conceptual level, the classification of Streng et al. (2009) is more concrete and application-related. The characteristics that they name are the *type of visualization*, the *placement* and the *type of information*. All these factors are possible characteristics of *mirroring systems*, *metacognitive tools* as well as *coaching systems*. However, Streng et al. (2009) mainly refer to *mirroring systems* in their work.

They define the *type of visualization* as either diagrammatic or metaphoric. Previous group mirror systems typically visualized information in a diagrammatic way, for example, in form of charts, such as bar charts or pie charts. An example is the SECOND MESSENGER of DiMicco et al. (2004) that uses bar charts in its first version. Another type of diagrammatic visualizations are matrices of light sources, used in the REFLECT table by Bachour et al. (2008). Streng et al. (2009) present the first metaphoric visualization for group mirrors. In a direct comparison of diagrammatic and metaphoric visualizations, they could show several advantages of metaphoric feedback compared to diagrammatic one. The second classification item is the *placement* of the group mirror on either table or wall. Streng et al. (2009) name pros and cons for both settings. With a table, information is in the center of the group, at the price of creating an orientation problem for some of the group members. Wall displays solve the orientation problem, but might not be in everybody's vision for groups seated around a table. The third classification part is the *type of information* that can either be quantitative or qualitative. The following design space adopts this differentiation, which will be explained in more detail in section 4.3.

4.2 Factors of the CSCW Design Space

Especially due to the interdisciplinary character of CSCW, there exist a large number of different classifications. One important and well-known example is the time and space matrix of Johansen (1988), another the CSCW design space of Mills (2003). Mills also refers to time and space as two of the ten key dimensions and extends them by the factors group size, interaction style, context, infrastructure, collaborator mobility, privacy, participant selection and extensibility. Ellis et al. (1991) present a taxonomy of groupware systems, focusing on the differentiation between types of systems such as message systems or computer conferencing systems. Carstensen and Schmidt (1999) distinguish tightly or loosely coupled interaction and they understand computers as medium or as regulator of interaction. In the following, I discuss the factors that are of specific interest for the design of group mirrors.

Space According to the time-space-matrix by Johansen (1988), collaborative work can take place either in the same space (co-located) or in different spaces (distributed/remote).

Both situations are possible usage scenarios for feedback systems and induce different affordances and challenges for the design of these systems.

Time The temporal aspect of the time-space-matrix illustrates that collaboration can take place at the same time (synchronous) or at different times (asynchronous). Transferred to the realm of group mirrors, this means that the feedback given through the system can either be shown in real-time or as a replay. Real-time feedback is more common in related previous work, but comes with a number of challenges, as the generation of the feedback content is time-sensitive. For replay feedback, it is possible to create more complex and thereby more time-consuming evaluations of the information, as the feedback will solely be provided to the group after the actual group work.

Task A basic factor in every kind of collaboration is the task that the group is working on. The task also plays an important role for the design of group mirror systems. The purpose of collaboration can be manifold, reaching from problem solving to idea finding to gaming. One possible categorization classifies these tasks from more structured ones to more open-ended ones with less predefined structures throughout the collaborative process.

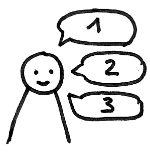
Group composition Several factors are of importance when talking about the composition of a group. One main factor is the size of the group, which can reach from a small team to a large group. Another aspect is the familiarity of the group members, how they are related to each other and if the group is structured in a hierarchical way or not. Furthermore, the expertise of the group members might vary, from novices to experts.

4.3 Factors of the Group Mirror Design Space

The most important variables of the design space for group mirrors are defined below. Counted among these are type of information, type of visualization, level of aggregation, feedback origin, feedback provider and receiver cardinality, level of anonymity of the feedback provider and receiver, placement and privacy, feedback type and amount of guidance.

Type of Information One differentiation made by Streng et al. (2009) is the classification of information into quantitative and qualitative types. Quantitative feedback captures quantitatively measurable information of collaboration such as speaking times, speaking turns, eye gaze or head orientation. This is usually done automatically by a system by using microphones to quantify speaking times or speaking turns or by using eye or head tracking systems to capture eye gaze or head orientation. In contrast, qualitative feedback gathers information about qualitative aspects of collaboration. Examples are the innovative strength of a contribution or the persuasiveness of an argument.

Quantitative information:

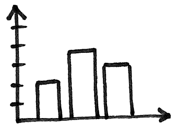


Qualitative information:

Type of Visualization There exist a number of taxonomies of visualizations, created by Lengler and Eppler (2007), Lohse et al. (1994) or Shneiderman (1996), for example. In the context of feedback systems for groups, Streng et al. (2009) differentiate between diagrammatic and metaphoric visualizations. A more general classification is a division into subtler and more overt forms. Schiavo et al. (2014) use a textual representation as an overt directive. Strictly speaking, textual representations are a possible representation for a group mirror. However, in the definition used in this thesis, group mirrors should convey the feedback in a subtle and unobtrusive manner. Thus, in our view, solely textual representations are not a manifestation of a group mirror. Nevertheless, the division into subtler and more overt forms is an adequate differentiation for group mirrors.

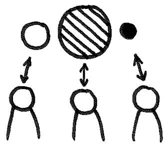
Subtle:
(e.g., metaphor)

Overt:
(e.g., diagram)

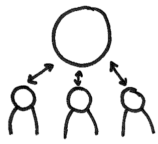


Level of Aggregation The information can either be visualized in an aggregated or individualized manner. An aggregated form of visualization shows the performance of the whole group while it is not possible to read off individual contributions. Accordingly, the contrary is a visualization that shows individual performance. This can imply that the feedback is more person-related or more task-related. Individual speaking time that is counted per person is an example for person-related feedback while the speaking time of all group members spent on a certain topic is an example for task-related feedback.

Individual:

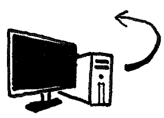


Aggregated:

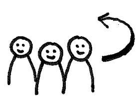


Feedback Creation The feedback that is provided to a group can be generated in different ways: automated through a system or via humans, be it experts or peers. Automatic systems are usually the best choice to collect quantitative information, but cannot trivially be extended to gather qualitative information. In the long run, machine intelligence is a promising approach to generating feedback on qualitative aspects. However, in the present state of the art, one fast and comparably reliable way of gathering qualitative feedback is using the intelligence of humans. The group members themselves can do this by providing peer feedback to each others. The second possibility is to involve people from outside of the group, such as independent observers.

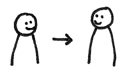
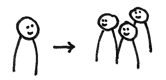
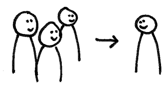
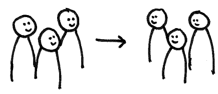
System:



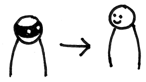
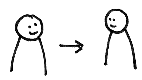
Humans:



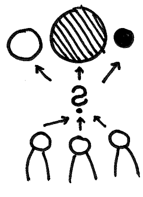
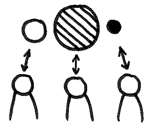
Feedback Provider and Receiver Cardinality There are different receiver cardinalities. The whole group may be the target audience for the feedback, or only part of the group. In the case that humans provide feedback, there are also different provider cardinalities. Taken together, this leads to the following possible provider and receiver cardinalities: *one to one*, *one to many*, *many to one* and *many to many*. *One to one* is a possible scenario in which, for example, in a teaching context, one person provides feedback to another person. In a *one to many* scenario, one person provides the feedback to the whole group. This is typically someone outside the group, for example, a moderator or an expert. In a *many to one* scenario, a group provides feedback only to one group member, a scenario that can be found in the classroom or in lectures where an audience provides feedback to the teacher or lecturer. A *many to many* situation is, for instance, a group in which each group member can provide feedback to the whole group.

One to one :

One to many:

Many to one:

Many to many:


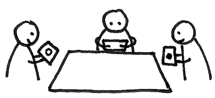
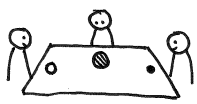
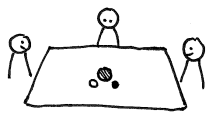
Anonymity of Feedback Provider In the case that humans provide the feedback, the originator can stay anonymous or be identifiable to the feedback receiver. An example of anonymous feedback in a *many to one* scenario is that the receiver can see the feedback of every individual without an indication of their identity. Accordingly, with identifiable feedback, the receiver can read off who has provided which feedback.

Anonymous:

Identifiable:


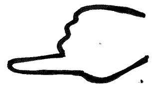
Anonymity of Feedback Receiver The feedback receiver can also be identifiable or stay anonymous. An example for identifiable feedback is a *many to many* scenario in which all group members may assign feedback to the whole group. The feedback is accessible to the whole group, it is for example, visualized on a shared display. Every group member has a distinct representation so that everyone can see the feedback for every other person. Contrarily, in the anonymous mode, all group members are represented with distinct units, though it is not possible to match the units to the individual participants.

Anonymous:

Identifiable:


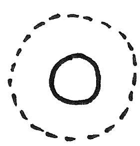
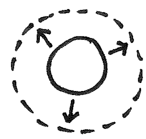
Placement and Privacy Placement and privacy are two interrelated aspects of group mirrors. Streng et al. (2009) differentiate between tabletop and wall displays. However, making a differentiation between public, semi-public and private feedback systems in the first place can broaden the aspect of the placement. Tabletop or wall displays can serve as public or semi-public displays. Public means that all feedback is visible equally well for each group member. Semi-public feedback is positioned in a way that individual feedback is best perceptible for the addressed group member and less perceptible for the others. Tablets or smartphones are examples for private displays. As feedback systems do not necessarily use displays to show the feedback, this classification is also applicable for other types of feedback. Auditory feedback, for instance, can be public by using one loudspeaker for the whole group or private by using headphones for each group member.

Private:  Semi-public:  Public: 

Feedback Modality Feedback can be received through the five senses and can therefore be visual, auditory, tactile, olfactory or gustatory. Visual feedback is most common in the group context. Besides feedback on displays, physical objects or light sources may be used to convey the feedback visually. Auditory or tactile feedback are also fitting choices. Auditory feedback can be more intrusive than visual feedback, as group members cannot simply look away and ignore the feedback. Tactile feedback, for example, in form of vibration, may be used to implement private feedback, as the receiver alone senses the feedback. Olfactory and gustatory feedback seem less appropriate and feasible in a collaborative situation, as they are difficult to distinguish and might be annoying for the participants.

Visual:  Auditory:  Tactile: 

Amount of Guidance Jermann et al. (2001) define three different levels of guidance for feedback systems: mirroring tools, meta-cognitive tools and guidance systems. While mirroring tools reflect the current state of the interaction, meta-cognitive tools provide a comparison between the current state of interaction to the desired state. Guiding systems additionally offer advice and guidance (for a more detailed explanation see Section 4.1). In the schematic pictures, the current state is represented with the continuous circle. The desired state is indicated with a dashed circle. Advice and guidance is represented with arrows.

Mirroring:  Metacognitive:  Guidance: 

Context of Use Another important aspect is context of use, meaning which goals are pursued by employing a group mirror. Concisely, these can be summarized as *learning how to conduct a certain collaborative technique*, *learning something about the subject of the collaboration* or *accomplishing the subject of the collaboration in the best way*. Group mirrors can support all these activities. To help to explain these three concepts, the example of argumentation can be adduced. Groups can *learn to argue*, *argue to learn* (for more work on this topic see e.g. Jonassen and Kim, 2010; von Aufschnaiter et al., 2008; Weinberger et al., 2010) or they can *argue for its own sake*, for instance, to come to a thoughtful decision. We did not incorporate this aspect in the schematic description, as most related work does not explicitly address this issue.

4.4 Classification of Previous Group Mirrors

A classification of group mirrors from related work is shown in Figure 4.1. A parallel coordinates visualization (Inselberg and Dimsdale, 1991) was used, with a slight alteration, which was inspired by the parallel sets visualization (Kosara et al., 2006). A parallel coordinates visualization is one of the standard methods in information visualizations to depict data in a multidimensional space. It consists of parallel axes and polylines with vertices on the axes. One line has one vertex on each axis. However, in the classification of group mirrors, one data point (a line in the parallel coordinates visualization) can have several vertices on one axis. Therefore, a variant based on the parallel sets visualization was used. This is a visualization for categorical data that depicts absolute frequencies (visualized in form of ribbons) on parallel axes. The ribbons can split up, a feature which I used for the visualization of the design space of group mirrors. This means that each line can split up and have several vertices on one axis.

I will explain this visualization in our use case and will give a concrete example from our data set. The different characteristics of the design space are displayed on the x-axis. The most important systems from related work are color-coded and listed at the top of the visualization. It can be read from top to bottom. To explain the visualization, I will briefly describe two examples. The MEETING MEDIATOR (Kim et al., 2008) is color-coded in red. The first axis is the characteristic of the **space** with the two manifestations *co-located* and *remote*. The MEETING MEDIATOR is designed for both scenarios, therefore, the red line splits up and reconnects in the next category, **time**, as the system is supposed to serve as a *real-time* support. The second example is the SECOND MESSENGER (DiMicco and Hollenbach, 2006), which exists in several versions. For the aspect of time, one version (the TIMELINE) is aimed at providing feedback as a replay, all others for real-time. Therefore, the line splits up, and the name of the version is written on top of the line.

It has to be noted, that this design space as well as the classification is not definite. For some of the categories, a continuum seems to be a more fitting choice than discrete points. Nevertheless, in order to give an overview of the design space, a simplified concretization was carried out. For each characteristic, a number of concrete manifestations were chosen,

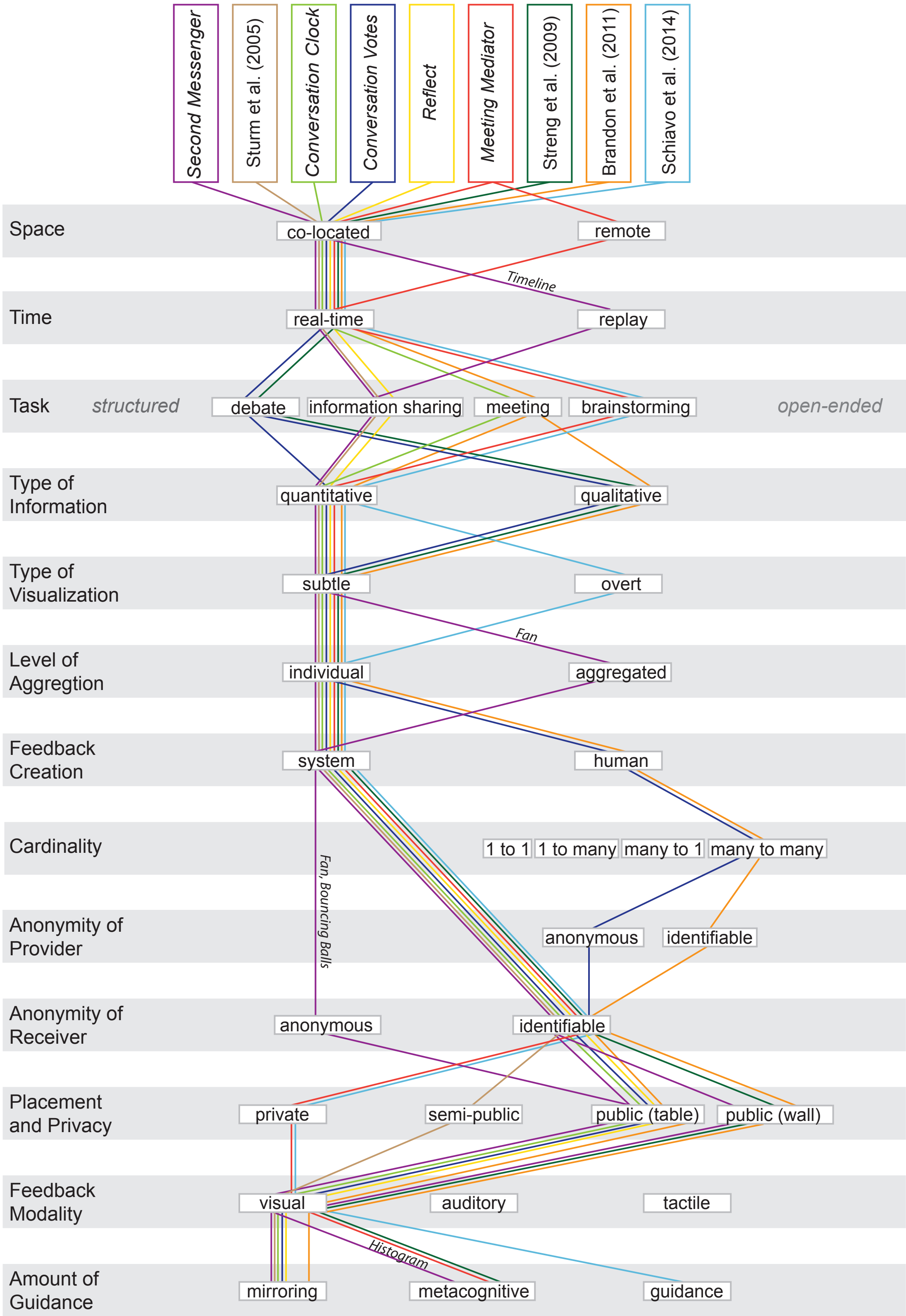


Figure 4.1: Design space. Co-located group mirrors classified according to the design space.

for instance, co-located and remote for the characteristic of space. For task, the scenarios were chosen that actually were used in at least one of the studies that were carried out with the systems. The data points for cardinality and anonymity of provider are included under the aspect human, as these only matter in cases in which humans provide the feedback.

In the following, the categorization undertaken in Figure 4.1 is described in more detail by discussing each characteristic, starting at the top. Explanations of some of the decisions for classifying the systems are outlined. A detailed description with exemplary images of the classified systems can be found in Section 2.2.2.

As described above, in regard to the aspect of **space**, two possibilities exist: co-located and remote. In this thesis, the focus lies on group mirrors for co-located collaboration. There also exist group mirrors for remote collaboration (often also called "awareness systems"). These are not depicted in this classification. However, one system, the MEETING MEDIATOR (Kim et al., 2008), is explicitly designed for co-located as well as remote collaboration and it is therefore categorized under both characteristics.

Switching to the next category, **time**, one can see that most of the group mirrors for co-located collaboration are designed for real-time deployment. Only one system is designed for the use as a replay. As depicted in the parallel coordinated diagram, this system is the TIMELINE version of the SECOND MESSENGER (DiMicco and Hollenbach, 2006).

For the category **task**, four different manifestations were chosen, namely the ones that were used in the user studies conducted with the systems. These are very structured tasks such as debates or (scripted) argumentation (summarized under "debate" in Figure 4.1), information sharing tasks, meetings or brainstorming sessions. In Figure 4.1, these tasks are ordered from more structured to more open ended.

The **type of information**, quantitative or qualitative, is ambiguous for two systems that can be classified under both categories. The CONVERSATION VOTES (Bergstrom and Karahalios, 2007b) system visualizes speaking times, a quantitative measurement, as well as voting, a qualitative aspect. The system of Brandon et al. (2011) combines quantitative feedback in form of connecting lines between speakers with qualitative feedback in form of agreement of group members.

Relying on Schiavo et al. (2014), a continuum from subtler to more overt was chosen for classifying the **type of visualization**. Only one version of a system from Schiavo et al. (2014) was categorized as overt, since it provided feedback in textual form.

The **level of aggregation** is individual for most systems, as the individual participants can be distinguished. Only the FAN version of the SECOND MESSENGER (DiMicco and Hollenbach, 2006) shows an aggregated visualization that makes a differentiation of group members impossible.

The **feedback creation** can either be executed by a system or it can be performed by a human. Systems that use a Wizard of Oz approach (see e.g., Kelley (1983)) in their study are classified under the category "system", as the Wizard of Oz technique presents a tool in a simulation of what may be possible in the future even without human intervention. Two systems actually involve human interaction in their design, the CONVERSATION VOTES system (Bergstrom and Karahalios, 2007b), in which group members cast votes on the discussion, and the system by Brandon et al. (2011) in which participants can indicate agreement or disagreement.

In the case that humans provide the feedback, the **cardinality** of the feedback provider comes into play. The cardinality can either be one to one, one to many, many to one or many to many. In the considered related work, only group mirrors realizing many to many feedback were implemented.

Another aspect is the **anonymity of the feedback provider**. He or she can either be anonymous or identifiable to the other group members. From the two systems that allow feedback from group members, one enables anonymous voting (CONVERSATION VOTES). In the other system (Brandon et al., 2011), others are able to identify who agrees to whom.

In both cases, when the feedback comes from a system or from a human, the **anonymity of the feedback receiver** can be determined. From the inspected systems, only the BOUNCING BALLS version of the SECOND MESSENGER enables the feedback receivers to stay anonymous while still representing each group member as an individual representation. The FAN is also classified under "anonymous", as in an aggregated visualization it is naturally impossible to identify individual persons.

The characteristics of **placement and privacy** are again more of a continuum. The MEETING MEDIATOR (Kim et al., 2008) and the tool by Schiavo et al. (2014) use private displays, however the other group members are aware of the information that is displayed on the private devices. The system by Sturm et al. (2006) displays feedback in front of the group members on a table, so that others can glance at feedback directed to the other participants.

In terms of the **feedback modality**, all of the group mirrors give the feedback visually.

The last category, **amount of guidance**, requires interpretation to some degree. Tools that provide an indication about an optimal state were classified as *metacognitive*, such as the HISTOGRAM version that indicates if group members are "over"- or "underparticipators". In case of the MEETING MEDIATOR one could argue that it is more a mirroring than a metacognitive tool, as the visualization does not provide an explicit notification about the optimal state. However, from another point of view, the optimal state is visually apparent as the chosen visualization implies that the circle should be aligned in the center of the screen. The same applies to the tool of Streng et al. (2009) in which the flourishing of the trees and the weather can be - borrowed from everyday language - "good" or "bad". The only system that was categorized as a tool providing guidance was the overt version of the system by Schiavo et al. (2014) that provides explicit hints on how to adapt the behavior.

4.5 Summary and Resume

The previous sections outlined a design space defining the most important characteristics of group mirrors. Related work was classified according to these characteristics and a variant of a parallel coordinates visualization was used to depict the focus of previous research on group mirrors for co-located collaboration.

The presented design space has three major benefits for future work. On the one hand, designers can make use of it when designing group mirrors. The brief overview of the important characteristics may be helpful especially in the phase in which initial design decisions have to be made. On the other hand, researchers have the possibility to use the design space as a support for their study designs. It is important to control the variables in an experiment and to understand the coherencies between them. When planning a study with a group mirror system, it may be helpful to verify by reference to the design space that only the intentionally chosen variables are altered. Moreover, it helps researchers interested in investigating the impact of different aspects of group mirrors to detect areas that already have been studied in detail and areas that need further evaluation.

4.6 Focus of the Present Thesis

In the scope of this thesis, due to the number of dimensions of the design space, I chose to make decisions regarding the aspects of the design space that should serve as dependent or independent variables for the studies conducted throughout this work. The reasons are outlined in the following.

First, this thesis investigates group mirrors for co-located collaboration that are applied in real-time. I assess that group mirrors for co-located collaboration are an hitherto less evaluated field than awareness systems for distributed collaboration. Further, we believe that it is important to provide supportive technology for co-located collaboration, as a trained moderator is not always available and especially novices in group work struggle in achieving successful and contenting collaboration. I further consider that a short feedback-loop of real-time feedback can lead to a faster amelioration of behavior and learning. However, these assumptions are subject to further investigation and are not explicitly addressed in this thesis.

Second, two tasks are investigated in more detail that cover the extremes of the category "task": professional debates are very structured tasks while creativity techniques are more open-ended tasks. That is why we take these two tasks as representatives for the breadth of the spectrum of possible tasks.

Third, especially qualitative forms of feedback are subject of the systems throughout this thesis. On the one hand, this area of research is underrepresented in related work until now. On the other hand, I believe that including qualitative feedback into group mirrors has the potential to further ameliorate group work. While currently, technological limitations make automatically provided qualitative feedback difficult, this might be possible in the future. Right now, an alternative to computer generated qualitative feedback is that humans provide this feedback. Providing feedback can be a worthwhile activity in itself and comes with a number of advantages when it is technologically-mediated. For instance, feedback from humans mediated through technology can be provided anonymously though all group members are situated in a face-to-face environment.

Last, all group mirrors presented in this thesis provide visual feedback. Reasons for this decision were that debates and creative collaboration are tasks with a focus on communication and conversation. We estimated that auditory feedback might interfere with these tasks. Due to the nature of the human senses, less information can be encoded and successfully distinguished with tactile feedback compared to visual feedback. However, tactile feedback could be a reasonable choice for real-time group mirrors, for example, in combination with visual feedback, and might be a fruitful area for future research.

III

SUPPORTING CREATIVITY WITH GROUP MIRRORS

5
Supporting Brainstorming

In the present work, I chose brainstorming (Osborn, 1953) as one collaborative creativity technique to investigate the influence of technologically mediated feedback in more detail. First, brainstorming is one of the most common creativity techniques. Though research has proven the advantages of electronic brainstorming systems (Connolly et al., 1990; Cooper et al., 1998), face-to-face brainstorming is still a common practice (Dennis and Reinicke, 2004). However, due to the problems of this form of brainstorming, additional support is especially necessary. Second, there is a large amount of research on brainstorming, which this work can build on and with which results of studies with group mirrors can be contrasted. In the following chapters, prototypes with the goal to support co-located brainstorming sessions are described and the results of several studies are outlined.

First, I will present GROUPGARDEN, a group mirror using metaphorical visualizations in form of a garden scenery, and two studies with this prototype. A general exploration of the concept and a comparison of two different display environments (table and wall display) will be outlined. Then, I will describe another prototype using the metaphor of balloons to investigate the issue of cooperative and competitive influences in brainstorming sessions. Finally, a similar system is used to evaluate another factor of the design space: the public visibility of the feedback.

This chapter is based on three bachelor theses (Kosan, 2013; Sachmann, 2014; Ta, 2014) and one project thesis (Raltchev, 2013). Part of it was published in the proceedings of two conferences (Tausch et al., 2014, 2016) with the co-authors Doris Hausen, Ismail Kosan, Andrey Raltchev, Stephanie Ta and Heinrich Hußmann. Some of the images and tables shown in this chapter have orignially been published in these papers. The detailed personal contribution statement can be found in the disclaimer.

5.1 Groupgarden: A Comparison of Table and Wall

In this section, I will introduce the prototype GROUPGARDEN, a group mirror designed to support co-located brainstorming sessions using metaphorical representations. After describing the motivation and outlining concept and design of the system, two studies are described. The preliminary study explores the general concept of GROUPGARDEN by comparing this system to a baseline without feedback. Building on the results of that study, the second study compares two versions of GROUPGARDEN, a table and wall version.

5.1.1 Background and Motivation

Our main goal with GROUPGARDEN was to develop a group mirror tailored to support face-to-face brainstorming sessions and to investigate whether such a subtle intervention can improve brainstorming and facilitate social processes. There already exist group mirrors, which have been applied in brainstorming sessions. However, these are not specifically designed to support brainstorming. As described in Section 3.1, the original form of brainstorming by Osborn (1953) is based on a number of rules. Furthermore, a number of problems have been described in research in the context of co-located brainstorming. Our prototype aims at supporting these rules and at defying the described problems. It has been reported that balancing participation in brainstorming is beneficial for both quantity and quality of ideas (Oxley et al., 1996). We therefore included a mechanism that is designed to make group members aware of the balance of contributions during the brainstorming session.

GROUPGARDEN's design uses metaphors derived from nature, in form of a garden scenery with flowers, a tree and the sky with clouds and the sun. This choice is inspired by other work, which I will outline in the following. The second study deals with the factor of the display environment (table or wall). I will also summarize related work from this area of research.

The Garden Metaphor

The main reason why we chose a metaphorical visualization for the design of GROUPGARDEN is a study conducted by Streng et al. (2009). This work is described in detail in Section 2.2.2, Figure 2.4 shows the metaphoric version during the study. The authors compared two versions of a group mirror with each other: a metaphoric and a diagrammatic visualization. The metaphors they used were trees and the weather, both representing the quality of contributions of different group members. The trees can change in gradation of five stages dependent on the quality of argumentation. These stages reach from a tree with leaves and fruits to a leafless tree. Similarly, the weather can change in five stages from cloudless sky to heavy rain. Furthermore, the times of day and night indicate different phases of the collaborative task. When group members interrupt each other, a lightning covers the display. The diagrammatic visualization provides the same information. A column chart in five stages is used to represent the quality of the group members' work. When collaborators interrupt each other, a red circle starts to flash. In a laboratory study, both versions were compared to each other. Results indicate that the metaphorical version supported self-regulative processes better than the diagram version and was further preferred by 70% of the participants.

Analogously to Streng et al. (2009), we relied on metaphors derived from nature because their effectiveness has been indicated in previous work. Furthermore, the metaphor of a garden is cross-culturally understandable and a relaxing and non-threatening environment, as Crossley et al. (1998) note. We presume that the organic nature of a garden is a feasible metaphor for representing the highly dynamic process of a collaborative creativity task. Specifically, we chose flowers and a tree for visualizing the most important aspects of collaborative brainstorming. Chau (2011) summarizes the advantages of the flower metaphor, which can be transferred to other plants such as trees as well. First, these metaphors are easy to understand (Lantin and Judelman, 2006; Xiong and Donath, 1999). Zhu (2002), for instance, could show that participants of her study showed a higher preference for the flower metaphor compared to a messenger without these metaphorical visualizations. Second, plants such as flowers and trees have a rich structure, which can represent different information that is not too difficult to be differentiated (Xiong and Donath, 1999). Third, positive and negative valuation can easily be encoded, for instance, by showing flourishing or withering plants.

We based the design of GROUPGARDEN not only on the work of Streng et al. (2009), but we were also inspired by other projects that used natural metaphors (see for example, Section 5.1). Another approach using the metaphor of flowers and a garden is the PEOPLEGARDEN by Xiong and Donath (1999). The concept of this work was already briefly described in Section 2.3.3. PEOPLEGARDEN uses so-called *data portraits* to represent individual participants of online interaction environments (e.g., chats). These *data portraits* are clustered to visualizations resembling a garden (see Figure 2.6, bottom left), enabling group members to see their individual data as well as providing them with the possibility to compare multiple group members. Xiong and Donath (1999) use the scenario of a Web-based message board. Flowers represent individual users. The petals of the flower represent the person's postings, arranged in a temporal order. The color of the petals shows if a post starts a new conversation (magenta) or if it is a response to an already existing topic (blue). On top of the petals, pistil-like circles indicate the number of responses from other users. Furthermore, the saturation of the petals gives an indication of the time of a posting, as the color of the petals fades with time. The height of a flower represents how long someone has been a member of the message board, deploying the metaphor that older flowers already grew taller. The metaphor of a "healthy garden" is used, as bright flowers indicate an active discussion.

Another example using flowers and trees as a metaphor for increasing motivation is the KNOWLEDGE GARDEN (Crossley et al., 1998). This is a 3D environment that uses the metaphor of flowers to represent information, in their case automatically clustered Internet resources such as bookmarks.

The ITREE project (Nakahara et al., 2005) is a mobile phone application with the goal to increase participation in online bulletin board systems (BBS). The flourishing of a tree and

Figure 5.1: Related work on metaphorical visualizations. The FlowerGarden by Lantin and Judelman (2006) (left), the UbiFit Garden by Consolvo et al. (2008b) (middle) and two sketches of a flower visualization by Polleti et al. (2012) (right).

the color of a sky change dependent on the number of posts, number of times posts are read, number of replies to posts and the ratio of forum posts to replies. The authors could show that the metaphors could increase participation in the online board.

Another example, the FlowerGarden (Lantin and Judelman, 2006), emerged from a collaboration of a designer and a computer scientist. Their concept comprises that each participant is represented by a flower with petals that indicate conversation events that they entered in an online tool (see Figure 5.1, left). Participants had some freedom in designing their flowers as they could choose between differently colored and shaped petals. A study *in the wild* revealed that participants competed for the biggest and pretties flower.

The previously discussed systems have in common that they want to motivate group members in some sort of collaborative process. Streng et al. (2009) aim at increasing quality of contributions in a scripted collaborative scenario, Xiong and Donath (1999) have the goal to enhance participation in online interaction environments, Crossley et al. (1998) designed their system for a collaborative 3D information visualization tool, Nakahara et al. (2005) intend at increasing participation in a BBS and Lantin and Judelman (2006) both present a collaborative development process and designed the tool for a collaborative environment.

Tools using metaphors derived from nature were also deployed in a number of other scenarios. The UbiFit Garden (Consolvo et al., 2008b) is designed for people who want to start including regular physical activities in their daily lives. Here, a garden scenery is shown on a glanceable display on a mobile phone (see Figure 5.1, middle). Butterflies represent goal attainments and different types of flowers represent different types of physical activities. In two studies (see also Consolvo et al., 2008a), the authors could show that participants especially perceived the glanceable display with the metaphors motivating. Results revealed that for participants without display, physical activity decreased over time and in phases in which maintaining physical activity is difficult (such as during holidays), which was not the case for participants that were made aware of their activities with the metaphorical visualization.

Another tool using natural metaphors aims at visualizing Web search results (Chau, 2011). "Flower glyphs" represent different information and meta-information of search results: The petals represent keywords, the length of the stem resembles the document length, the number of leaves indicate external and internal outlinks and the size of the ground the flower is standing on represents the number of inlinks.

Lastly, I will briefly refer to a tool designed to lead to a continued engagement, in their example by reducing one's personal carbon footprint (Polleti et al., 2012). Therefore, different design approaches were developed: visualizations using flowers (see Figure 5.1, right), pie charts, jelly fish and footprints. In summary, these studies and projects show that metaphors have found their way into a number of different technological tools with the goal to engage people in certain tasks. Usage scenarios reach from collaboration tools to systems designed for more engagement in physical activities. Several studies could show the benefits of these kinds of visualizations. However, the influence in a brainstorming scenario has not been investigated yet. A first prototype for this scenario is described in this section.

Comparisons of Table and Wall Displays

Another aspect that the study with GROUPGARDEN investigates is a comparison of table and wall displays. Evaluating the advantages of various types and sizes of displays has been a topic of research since these possibilities exist.

Mandryk et al. (2002) discuss seven different display factors, among them the *orientation*, the *privacy* and the *number of displays*. The authors review the impact of these factors on group work. With the orientation of displays, Mandryk et al. (2002) refer to vertical (walls) and horizontal displays (tables). The main distinctions are that vertical displays offer the same view for all group members while people sitting around a table have different views. This can have an impact on several factors of the mechanics of collaboration described by Gutwin and Greenberg (2000), namely on coordinating, planning and monitoring of tasks. Besides that, particular objects may orient differently to the borders. On walls, objects frequently align with the lower edge while on horizontal surfaces, objects often do not align to the surface borders. Also, transient objects (such as mugs or notepads) can easily be placed on horizontal surfaces.

Inkpen et al. (2005) present a number of field studies investigating among other aspects the *display angle* (referred to as *display orientation* by Mandryk et al. (2002)). They found differences regarding communication. Participants made more pointing gestures at a table display compared to a wall display. On the one hand, participants commented that sitting around a horizontal surface was more comfortable and natural than working with a vertical display. On the other hand, participants also reported that they were more focused on the task when working on the wall display. Other differences were found in regard to ergonomics. On the wall display, people tended to write larger and some of the participants stated that writing was more difficult. Furthermore, people tended to move more when working with a wall display.

Another direct comparison of horizontal and vertical displays was conducted by Rogers and Lindley (2004). Their results show that working on a tabletop display led to a more cohesive working style, while vertical displays are perceived as less beneficial for group work. However, their study only covered small groups, in their case, a group with three group members. For larger groups, vertical displays have a number of advantages, as more people can gather in front of a wall display compared to a tabletop display. Moreover, the task they chose required group members to create a number of different representations, meaning that interacting with the surface played an important role for the task. A comparison of table and wall displays in the specific context of a scripted collaboration has been conducted by Streng (2010). This study is described more in-depth in section 3.2.3. Figure 3.4 shows the setting of the system on table and wall displays.

The before described investigations of horizontal and vertical displays have in common that they anticipate displays with which group members can interact directly, for example, by touch or multi-touch gestures. Group mirrors, however, do not allow direct interaction as they are designed to stay in the periphery of the group's attention. Considering that group members can influence the visual representation more indirectly by adapting their behavior, the focus of the study that is described in the course of this section focuses on the effects of horizontal and vertical displays that do not allow direct interaction.

5.1.2 Concept and Design

GROUPGARDEN consists of two interfaces: (1) a *feedback interface* that displays the feedback to the group, which is projected on a large surface and (2) a *control interface* that runs on a regular computer and that is operated by a person outside of the group. Both interfaces are implemented with Adobe Flash[1] using ActionScript 3.0[2]. A client-server architecture is used to enable the two interfaces to communicate with each other (for a detailed description see e.g., Winter, 2011).

The *feedback interface* is designed to support the brainstorming rules and to diminish possible problems. The probably most influential rule is to focus on quantity rather than on quality of ideas. To facilitate group members to find as many ideas as possible, the number of ideas of each group member is visualized. The metaphor of flowers is used to accomplish this. Figure 5.2 shows a sketch of the *feedback interface* of GROUPGARDEN. A schematic description of the main functionality is drawn in in purple. Flowers represent the individual group members. In the depicted example the group consists of three participants. The number of petals as well as the height of the flower represent the amount of ideas. At the beginning, all flowers are equally high (level 1) and all petals are unfilled (visualized with dashed outlines). With each idea of a group member, one petal of the flower that represents

[1] http://www.adobe.com/products/flashruntimes.html, last accessed on 26.04.2016

[2] http://help.adobe.com/en_US/FlashPlatform/reference/actionscript/3/index.html, last accessed on 26.04.2016

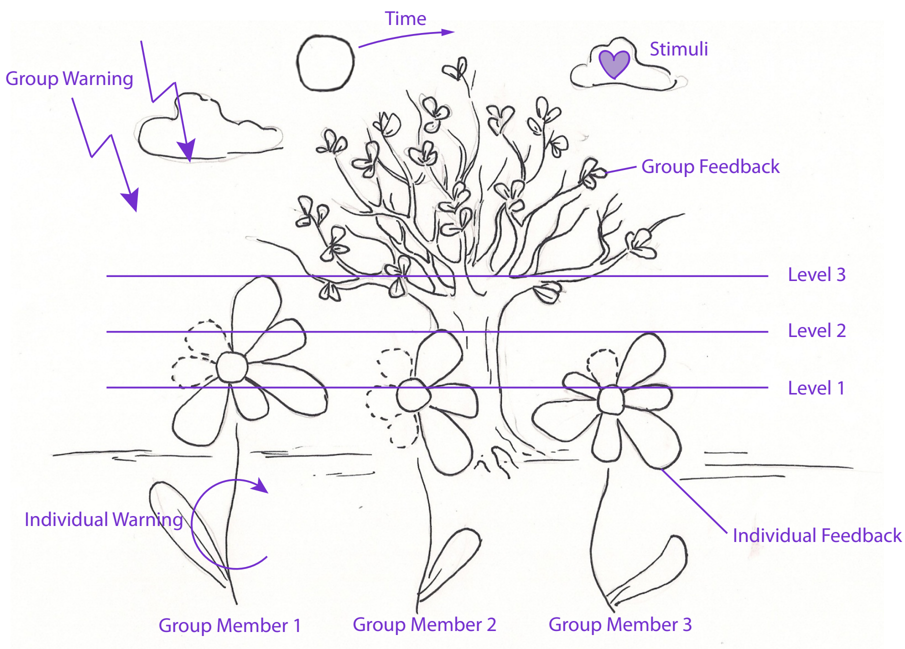


Figure 5.2: Concept of GROUPGARDEN. An annotated sketch of the concept.

the group member fills up. When all petals are filled, the flower grows (meaning that its height reaches level 2). New, unfilled petals appear.

Another mechanism that should motivate groups to produce more ideas is to provide feedback about the group's performance in addition to the individual feedback. At the same time, this aims at balancing participation, which has proven to be beneficial for brainstorming (Oxley et al., 1996). In our prototype, the metaphor of a tree is used to visualize group performance. In the beginning, the tree has no leaves. Every time when all flowers reach a certain level, more leaves and fruit grow. This means that the minimum mutual level of the flowers determines the appearance of the tree. For example, when all flowers reach a minimum level of 2 (one or more flowers can already have reached higher levels), the tree will start to flourish.

To our knowledge, a comparable combination of individual and aggregated feedback has not been implemented in previous group mirrors. The additional feedback about the group aims at generating a positive group experience and strengthening the common goal. Furthermore, the use of gaming elements is intended to further increase participation. Gamification is the *"use of design elements characteristic for games in none-game contexts"* (Deterding et al., 2011). In our case, gaming elements are the avatars in forms of flowers and trees that represent individual group members and the whole group. Comparable to a game, group members should strive to collect as many petals as possible and at the same time pay attention to the tree, which only flourishes when everyone contributes in a fairly balanced way.

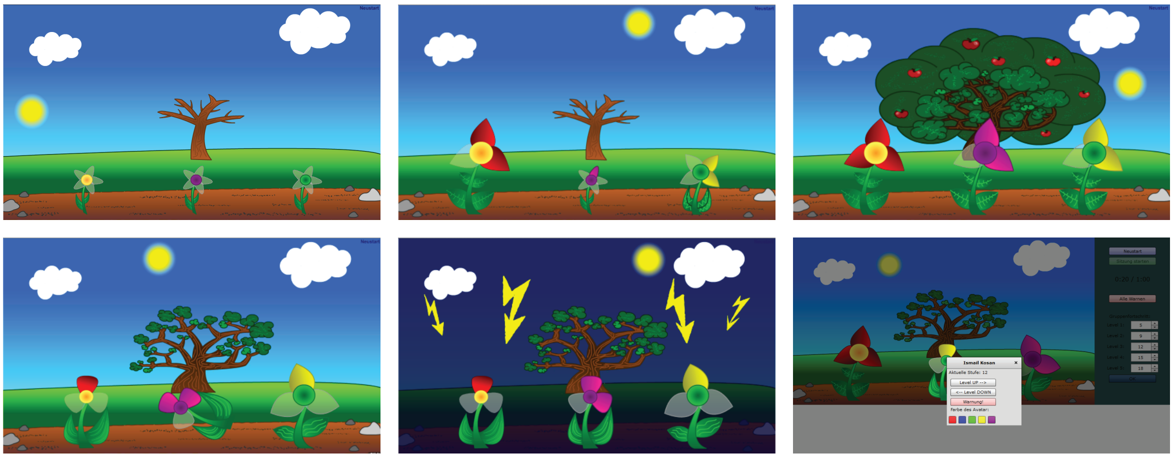


Figure 5.3: GROUPGARDEN Visualizations. Top row: GROUPGARDEN in the beginning (left), an extremely unbalanced brainstorming session (middle), and a balanced brainstorming session with a lot of ideas (right). Lower row: Individual warning in form of a rotation of a flower (left), group warning in form of a lightning (middle), and the *control interface* (right).

The third mechanism designed to increase the number of ideas is to show stimulating pictures in phases of silence. This idea is based on a project by Wang et al. (2010, 2011), who could show the positive influence of conversationally retrieved pictures on brainstorming. With this, the flow of ideas should be reflated when motivation decreases or the flow of ideas comes to a standstill. Furthermore, these pictures can be shown not only in phases of silence but also when a group gets stuck in a certain direction of ideas. Random pictures appear in the clouds in those cases. At the same time, this is thought to lead to more wild and unusual ideas, as the pictures are not necessarily related to the topic of the brainstorming.

Two other rules that our system aims to support are that criticism is not permitted and that group members should not interrupt each other. In both cases, individual group members need to be addressed. In these cases, the appearance of the flower changes by rotating the whole flower (see “Individual Warning” in Figure 5.2). Only one type of warning is used as we assume that group members know if they criticized an idea or if they interrupted another person. We deliberately chose a noticeable rather than a more “natural behavior”, as subtle changes on a peripheral display easily remain unnoticed.

Finally, the group should stay on topic. A group warning is used to make deviations from the topic apparent. This is, in reference to the work of Streng et al. (2009), done by showing a lightning (see “Group Warning” in Figure 5.2). Additionally, the metaphor of the sun that moves over the sky in a semicircle from left to right represents the passed time and the time that is still scheduled.

Figure 5.3 shows the final design of GROUPGARDEN. The image in the upper left displays how the visualization looks in the beginning: All flowers are on level 1, the petals are unfilled, the tree does not have any leaves and the sun is on the left of the display. The image in the middle of the top row shows an extremely unbalanced brainstorming. The person represented by the flower on the left has an already large flower while the group member represented by the flower in the middle is still on the first level. Thus, the tree did not start to grow. The image on the upper right shows a balanced brainstorming session with a lot of ideas produced by all three group members. The lower row shows how individual warnings in form of rotating flowers look like (left) and how group warnings in form of a lightning look like (middle.)

The *control interface* is displayed in Figure 5.3 at the bottom on the right. A person outside of the group operates this interface that then updates the visualization that is displayed to the group. The possible functionalities in the menu on the right are to start and end the session, to set a time that is scheduled for the brainstorming session and to trigger group warnings. By clicking on a flower, a pop-up window opens on top of the flower. There, it is possible to individualize the flowers by choosing a color. Furthermore, the counter for the number of ideas (i.e., the number of filled petals) can be increased. Decreasing is also an option, primarily for undoing a mistakenly increased idea counter. In this window, individual warnings can be triggered.

5.1.3 Preliminary Study: Evaluation of the Prototype

We conducted a preliminary study to investigate the concept of GROUPGARDEN. For this, we compared the prototype with the baseline without supportive feedback. The main research questions of this study were, whether GROUPGARDEN facilitates the ability of self-regulation of participants and whether brainstorming rules are successfully supported.

Method

The experiment was run as a laboratory study using a repeated measures design. This means that all groups accomplished two brainstorming sessions, one with each condition. One condition served as a baseline. In that condition, groups brainstormed without any additional support. In the other condition, feedback was provided with GROUPGARDEN. We used two different topics. Both conditions and both topics were counterbalanced using a Latin square design.

Setup and Procedure

The study took part in a quiet room. Ten groups with three participants each took part. The room was equipped with three revolving chairs and a projector that was used to project the visualization onto a white wall (see Figure 5.4, left). Participants could choose how to position themselves in front of the wall (more side-by-side or more face-to-face) (see Figure 5.4, right).

Before the sessions, the experimenter gave an introduction about the procedure of the study. The brainstorming rules were explained to the groups. Groups were asked to try to follow these rules. They were not explicitly asked to strive for balanced participation. Before the

Figure 5.4: Study setup. Left: A group of three participants in front of GROUPGARDEN (picture re-staged). Top right: A group sitting more face-to-face to each other. Bottom right: A group sitting more side-by-side. The visualization is projected on the wall left of the participants (not visible in these pictures).

condition with group mirror, its functioning was explained. Furthermore, the topics were given to the groups. We chose two topics suited for brainstorming that did not require any special precognition: (1) *What could a commercial for a new tablet computer look like?* and (2) *What could a commercial for a new caffeinated soft drink look like?*

After the general introduction groups brainstormed twice for 15 minutes. The brainstorming was accomplished in a purely verbal form without taking notes or using other stationery tools, such as sticky notes. The study was designed as a Wizard of Oz experiment (Kelley, 1983). This means that participants of such a study think that they interact with an autonomous system, while this is actually, at least partly, operated by humans. In our case, participants were told that the experimenter only takes notes on a laptop while actually he additionally operated the *control interface*, meaning that he increased the idea counter or triggered warnings.

The experimenter had to listen to the discussion carefully to estimate, which contributions should be counted as an idea. To standardize this procedure, we defined what should be counted as an idea. As described in Section 3.1, definitions of creativity often include the aspects of *novelty* and *usefulness*. The definition used for this study focuses on the aspect of *novelty*: Each contribution that is on-topic and is novel in the context of this brainstorming session (i.e., was not stated before) is counted as an idea. Additionally, ideas building on the ideas of others need to include a somehow novel facet to be counted as idea. We did not include the aspect of *usefulness*, as we suppose that it is too difficult for the experimenter to estimate the usefulness of an idea in real-time during the discussion.

To estimate the reliability of this real-time coding of ideas, two persons (the experimenter and another person) coded the two brainstorming sessions of the first group using the video recordings. The coders were not allowed to pause and replay the video, as the situation should be as similar as possible to the real-time coding scenario during the study. Cohen's kappa showed substantial agreement between the two coders (κ = .80).

After each condition, participants filled out pen-and-paper questionnaires with 5-point Likert scales (1 = strongly disagree, 5 = strongly agree). At the end, participants additionally filled in a final questionnaire asking about perceived differences between the two conditions, and about preferences and demographic information. A short semi-structured interview with the whole group was held and all participants were debriefed.

All sessions were audio and video recorded. Videos were taken so that all group members are visible from the front. A screen capture was taken from the control interface that was synchronized with the videos afterwards.

Participants

In the experiment, 30 voluntary participants took part in groups of three (12 female; average age: 24, range: 18 to 32 years), 22 of them were students, 6 research assistants, 2 stated other professions. In eight groups, participants already knew each other before the study. Participants could choose if they receive a 10€ voucher from a well-known online store or participate in the study as part of an obligation in their study program.

Results

The study was evaluated using the questionnaires, interviews, video recordings, loggings and the notes of the experimenter. A dependent t-test was used to evaluate quantitative information and a Wilcoxon Signed-Rank Test for evaluating the results from the questionnaires. A 5% level of significance was applied for the tests. We used Excel for calculating the t-tests and the statistical software SPSS for calculating the Wilcoxon Signed-Rank Test.

Performance The mean number of ideas per group was 38.2 in the baseline (SE = 3.2) and slightly more, 40.9 ideas, in the group mirror condition (SE = 1.3). A t-test did not show a significant difference. The concepts that aimed to increase the amount of ideas were, as described in Section 5.1.2, the use of individual representations in form of flowers, a group representation in form of a tree and pictures shown in the clouds.

To understand the influence of these factors on the results better, we evaluated how often and with which reasons images were displayed. In three groups, there was no necessity to show any images. In one group, an image appeared once, in another group twice. In three groups, three images appeared and in two groups, four images appeared during the brainstorming. However, the reason for showing the images was rarely that the idea flow came to a standstill, but was mostly to encourage groups to think also in other directions and to include more wild ideas. This indicates that in our study, the images might have had a minor impact on the number of ideas.

Despite the little difference in the number of ideas between the two conditions, participants still perceived themselves and their group members as more productive with support of the

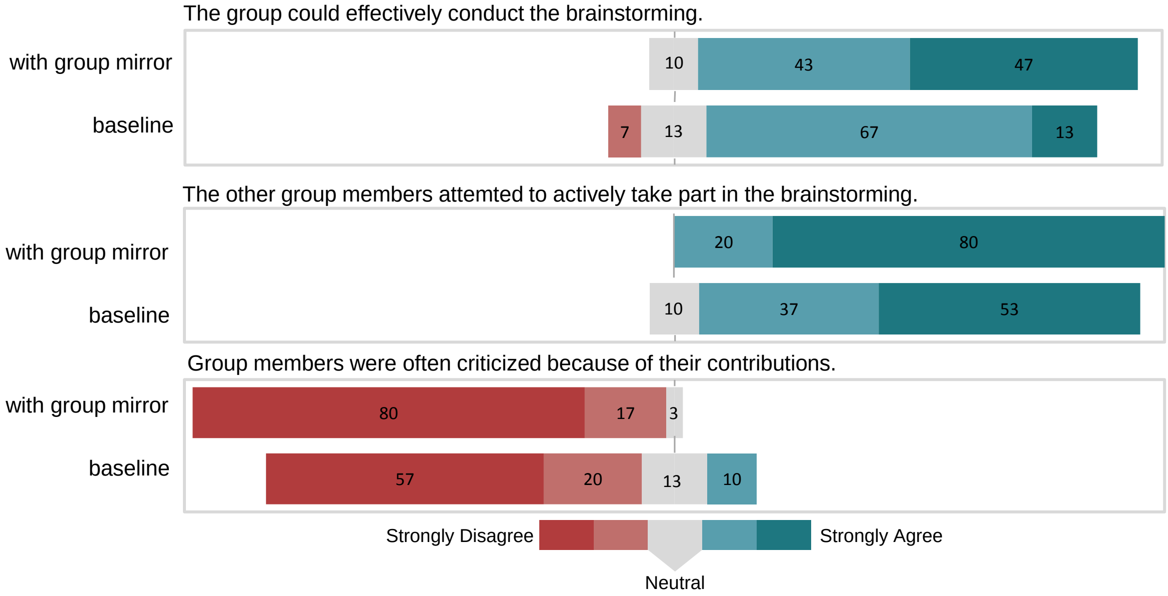


Figure 5.5: Results of the questionnaires. Questionnaires handed out after both conditions (with group mirror and baseline). Numbers indicate the percentage of participants who answered with that score on the 5-Point Likert skale.

group mirror. In the final questionnaires, 91% of the participants stated that they were more motivated in the group mirror condition than in the baseline condition. Comparing the questionnaires that were handed out after both conditions revealed that participants perceived the effectiveness of the brainstorming session with group mirror better than without (z = -2.37, $p < .05$, r= -.53), (see Figure 5.5, diagram at the top). They also rated the effort of others to participate in the brainstorming better in the group mirror condition compared to the baseline (z = -2.67, $p < .05$, r = -.6) (see Figure 5.5, diagram in the middle). For instance, one participant stated: *"I found that the second time [baseline] we somehow again and again discussed the same topics. I think the first time [group mirror condition] we have more productively addressed new ideas because you wanted to get bigger [flowers]."* (G6, P3).

Balance of Participation To assess the balance of the amount of the ideas of the group members, we categorized participants into *below average* and *above average*. This categorization was done after the study. Thus, group members did not get to know their categorization. To realize this, we took the baseline as basis and divided the mean number of ideas per group by three to get the mean number of ideas per participant (M = 12.7). All participants with more than this amount of ideas (i.e., at least 13 ideas) were categorized as *above average*, the others as *below average*. This resulted in 17 *above average* and 13 *below average* participants. We calculated a dependent t-test for both of these groups. Results show that *above average* participants contributed significantly less ideas in the group mirror condition (M = 13.76, SE = .48) compared to the baseline (M = 16, SE = .74), $t(16)$ = 3.27; $p < .005$; r = .63. *Below average* group members in contrary participated significantly more

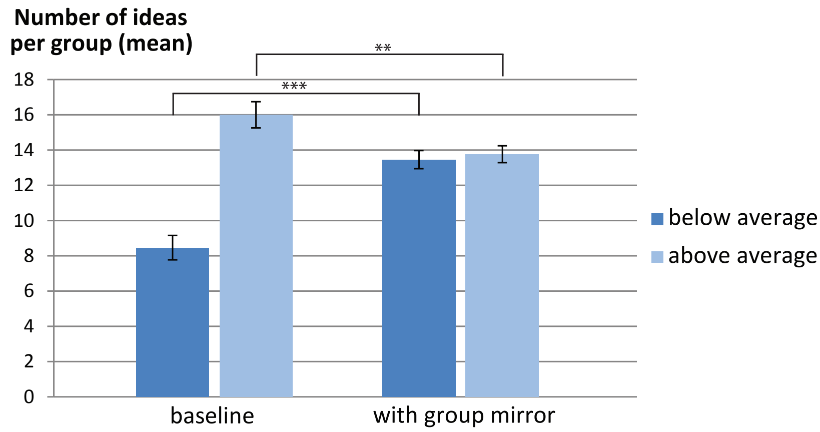


Figure 5.6: Results on the number of ideas. *Below average* participants contributed more with group mirror compared to the baseline, *above average* participants less, leading to a more balanced brainstorming session. Error bars represent the standard error.

with group mirror (M = 13.46, SE = .51) compared to the baseline (M = 8.46, SE = .69), $t(12)$ = -5.36, p < .0001, r = .84. Figure 5.6 visually shows this effect.

These results are also supported by the answers from the questionnaires and interviews. 73% of the participants perceived participation levels as more balanced with support of the group mirror compared to the condition without that support. In the interviews, both *below average* and *above average* participants stated that they altered their behavior when supported with GROUPGARDEN: *"I didn't want to be the one with the ugliest flower and bugger up the growth of the tree."* (G1, P1) and *"You restrain yourself more if you see that your flower is already bigger. I stopped talking then and thought: 'let the others talk'"* (G4, P2).

At the same time, participants did not feel obliged through the group mirror to balance their participation levels at any cost. Participants stated in the interviews that *"the system would not restrain me from saying something, if I had a really good idea"* (G1, P1) and that *"if the others don't come up with ideas at that moment I would still go on talking because then again you inspire the others."* (G1, P2).

Interruptions As stated before, we were interested whether individual warnings (that were provided in case of interruptions or judgments of ideas) had an effect on groups. Hence, we measured the number of interruptions and judgments. All interruptions or cases of simultaneous speech were part of the natural conversation. For example, when two persons started to speak at the same time, we did not count them as avoidable occurrences of interruptions.

Judgments Individual warnings in case of judgments were also observed rarely, however, significantly more often in the baseline compared to the group mirror condition. With group mirror, three occurrences of judgments were noticed and thus a warning was displayed (M = .03, SE = .21). In the baseline, group members judged ideas of others 15 times (M = 1.5, SE = .34), $t(9)$ = 4.81, p < .001, r = .85. This was also observed by the participants, who answered in the questionnaires that more occasions of judgments occurred in the

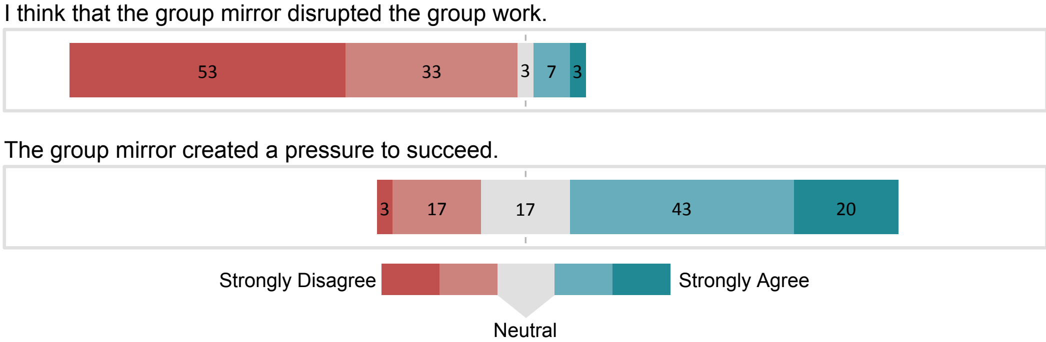


Figure 5.7: Results of the questionnaires. Questionnaires were handed out after the group mirror condition. Numbers indicate the percentage of participants who answered with that score on the 5-Point Likert scale. Numbers are rounded and thus might not add up to 100% exactly.

baseline condition compared to the group mirror condition (z = -2.62, $p < .01$, r= -.59) (see Figure 5.5, diagram at the bottom). These results are also reflected in this statement from the interviews: *"With the feedback system I took care to completely leave out any criticism."* (G4, P1).

Deviation from the Topic Group warnings in form of a lightning were also needed rarely. In the group mirror condition, four groups were warned once about digressing from the topic (M = .04, SE = .16), in the baseline, deviations from the topic were recorded seven times (thereof three times in one group). A comparison between both conditions did not reveal a significant difference.

Distraction and Pressure In the questionnaires, we asked participants, if GROUPGARDEN distracted them and if they felt under pressure. Results indicate that the group mirror is little distracting, but created a pressure to succeed (see Figure 5.7).

In the final questionnaires, one participant wrote: *"It is difficult for me to say, if the negative effect of the pressure the feedback produces outweighs the advantages that come from this pressure, to generate ideas by all means."* (G1, P2). Another participants stated: *"I perceive it as rather disrupting to constantly see the progress of the brainstorming, because the feeling emerges that some kind of pressure is produced."* (G4, P1)

Seating Arrangement Groups were free to choose their seating arrangement. Six groups positioned themselves in form of a triangle in front of the visualization. This induces a situation in which all group members can see each other, but for two of them, the visualization is in the periphery of their field of vision. The remaining four groups sat in a row, such as in a cinema, thus, making eye contact more difficult but having a good view on the group mirror. Both seating arrangements were perceived as not ideal. Participants remarked: *"I found the seating arrangement a bit difficult. Now [in the baseline] I found it much more pleasant that we could sit in a circle and look at each other."* (G2, P2). *"It would be cool if we could have the visualization more centered, or on all walls."* (G9, P2).

Preferences and Design When asked, which brainstorming session participants liked more, 74% answered in favor of the session supported by the group mirror. Furthermore, 70% were more satisfied with the results of the brainstorming. As already stated above, 91% felt more motivated, two participants explicitly mentioned that the "playfulness" of GROUPGARDEN was the main reason. One participant stated: *"This is like a task for all of us, like a game"* (G6, P2).

Most participants liked the visual design of GROUPGARDEN. When asked in the questionnaires whether they liked the visual representation, 93% either fully agreed or agreed, 7% stated that they did not have an opinion. The visualization was rated as intuitive and simple. *"You know immediately what it means."* (G1, P2). However, it was also perceived as *"(...) still designed childlike."* (G4, P2). It was remarked that it depends on the usage scenario, if the design is appropriate: *"(...) I wouldn't give it to a businessman, but for us it was actually pleasing."* (G5, P3). An aspect that was confusing and was not well received were the images that were displayed in the clouds.

Summary and Discussion

With GROUPGARDEN, we presented a group mirror that displays feedback about behavior during a brainstorming session to a group. It combines individual and group feedback to increase the number of ideas and to balance participation.

However, the amount of ideas did not differ significantly between the group mirror condition and the baseline. A possible explanation for this is that phases of silence occurred rarely. Groups did not come to a point where they ran out of ideas. This indicates that in phases of constant idea flow, as in our study, people do not create more ideas with support by a group mirror than without that support. At the same time, this indicates that the time spent on explaining individual ideas seems not to decrease with group mirror. However, we assume that without a temporal limitation (in our study, each brainstorming session was restrained to 15 minutes) the group mirror might have a positive impact on the quantity of ideas. When groups run out of ideas in the end of a brainstorming session, the group mirror might encourage them to continue and think in other directions. A general statement about the issues about achieving significant results in the studies presented in this thesis is provided in Section 9.2.5.

GROUPGARDEN successfully helped to balance participation of group members. However, this means at the same time that very productive group members decreased their amount of contributions. As the overall amount of ideas did not decrease, this can be seen as a positive effect, as it shows that free riding is less likely and group members who tend not to contribute are motivated to participate more actively.

Next to increasing quantity of ideas and balancing participation, GROUPGARDEN aims at minimizing occurrences of interruptions, judgments and deviations from the topic. Results showed that ideas of others were judged significantly more often in the baseline condition. However, all of these issues were noticed rarely. This might be due to the artificial situation of the laboratory study. Participants were aware that they were observed and recorded. In more natural setting, these issues might occur more often and therefore the group mirror might have a bigger impact on these issues.

We observed that the usage scenario of such a feedback system is an important factor. The design of GROUPGARDEN was considered more appropriate for informal use cases than for professional contexts. Creating different visual designs for different use cases can solve this issue. After having conducted this study and after having observed groups using this system, we assume that GROUPGRADEN is more appropriate for learning how to brainstorm, than to use it as a constant support. This would mean that groups use the group mirror until they have internalized the rules of brainstorming. More unobtrusive designs that demand less attention might be more appropriate using them as a constant support during group work.

Finally, we observed that displaying feedback on a wall led to several problems. Groups positioned themselves differently in front of the wall. When sitting more side-by-side, eye contact and a natural conversation was complicated, when facing each other, the visualization was in the periphery of the vision field for several group members. Therefore, we decided to study the effects of the display setting in more detail. The adaption of the initial prototype and the study comparing two different display settings will be explained in the next sections.

5.1.4 Adaption of the Initial Prototype: A Tabletop Version

Several participants of the preliminary study were dissatisfied with the position of the group mirror. We therefore decided to build a version of GROUPGARDEN that can be displayed in a more centralized style, namely on a table. Related research (see Section 5.1.1) has investigated the effects of different display environments on group work and could show that this aspect can have a not to be neglected impact. However, related work mainly focused on interactive surfaces, while group mirrors only allow indirect interaction. We therefore decided to conduct a study to compare the influence of the display setting (table or wall) of group mirrors on collaborative processes.

We decided to use the initial version of GROUPGARDEN for the wall condition without changing it much. We only removed the functionality of the images that appeared in the cloud as these were perceived as confusing in the first study. Our goal for the second study was to create a tabletop visualization that is as similar as possible to the wall version while taking the advantages of a tabletop display.

The tabletop version of GROUPGARDEN shows the same scene as the wall version: a garden with flowers and a tree. But in this version the beholder looks at the scenery from an aerial view (see Figure 5.8). The flowers and trees can now be seen from the top. To maintain the color scheme, the garden is surrounded by water instead of the blue sky. To project the visualization on a table, a mirror is attached to a video projector in an angle of 45 degrees.

This setup makes the assignment of the avatars to the group members easier. However, the metaphor of the sun does not work equally well in this version, as in an aerial view of such a

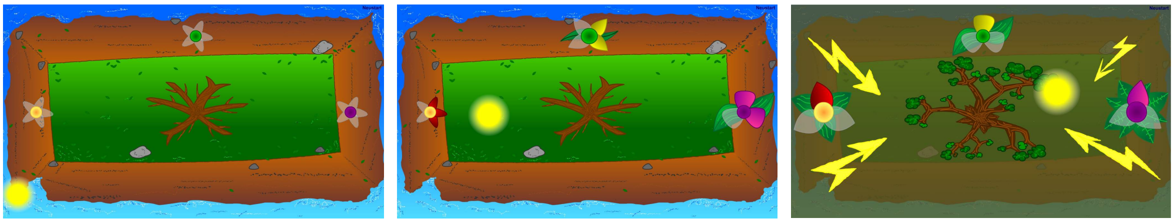


Figure 5.8: The tabletop version of GROUPGARDEN. From left to right: Scenery in the beginning of the brainstorming (left), after some ideas have been generated (middle) and a situation in which a group warning appears (right).

scenery, the sun would still be above the beholder. Solutions for better suited visualizations are possible (e.g., moving shadows of the flowers and the tree), however, due to reasons of comparability, we decided to keep the metaphor of the sun as a timer.

5.1.5 A Comparison of Table and Wall Display

The second study with GROUPGARDEN is designed to compare the wall and the table version. The main questions are, whether the location influences self-regulation of the group, whether brainstorming rules are better supported by one of the two display settings and whether it is possible to find out differences between both.

Method

Similar to the first study, the experiment was designed as a laboratory study using a within groups design. Again, two conditions were compared to each other, in this case, the wall and the table version. The same two brainstorming topics were chosen. Again, conditions and topics were counterbalanced.

Setup and Procedure

The main setup and procedure of the study was equal to the first study (see Section 5.1.3). The setup differed only in one aspect. This time, the seating arrangement was predefined. A square table was positioned in the middle of the participants. In the tabletop condition, the group mirror was projected on the table, in the wall version, the table stayed empty (see Figure 5.9). The procedure also differed only in one aspect. In the present study, the group mirror visualizations were explained before each condition, in contrast to the first study, in which no visualization had to be explained before the baseline condition.

Participants

Here too, three participants built one group. None of the participants had taken part in previous GROUPGARDEN evaluations. We conducted 8 sessions with 24 participants in total (11 female; average age: 25 years, range: 20 to 34 years), 18 of them were students, 3

Figure 5.9: Study setup. Left: Tabletop condition. Right: Wall condition.

research assistants, 3 stated other professions. In one group, all group members knew each other before the study, in 12 groups, two of the three group members knew each other and in the other 11 groups, participants did not know each other. Again, participants could choose between a 10€ voucher or credits for their studies.

Results

The captured data (responses from questionnaires on five-point Likert scales and interviews, video data, loggings and notes of the experimenter) were evaluated using the Wilcoxon Signed-Rank Test for Likert scale items and dependent t-tests for all other quantitative data. A 5% level of significance was applied. The statistical software SPSS was used for calculating the Wilcoxon Signed-Rank Test, Excel for calculating t-tests.

Performance Results of the second study do not reveal a significant difference regarding the amount of ideas between the wall condition (M = 37.63, SE = 3.47) and the table condition (M = 39, SE = 3.77). This is also supported by the answers of the questionnaires. In the final questionnaire, we asked if participants were more motivated in the table setup. 37% strongly agreed or agreed, 17% were neutral and 46% strongly disagreed or disagreed. In addition, we asked, in which condition participants were more motivated (providing only the two conditions as possible choices). 42% felt more motivated in the table condition, 33% in the wall condition and 25% did not decide and ticked both options. This indetermination is also reflected in some comments from the interviews. So, one participant stated about the table version: *"The group was rather more closed and the collaboration was a bit better, but [I was] only a bit more motivated, because I was motivated with both systems."* (G1, P1).

Balance of Participation To evaluate, whether participation was more balanced in one of the two conditions, we classified participants in *above average* and *below average* participants, as in the first study. This time, we took the wall condition as basis. The mean number of ideas in this condition was 12.54, resulting in a categorization of 13 *above average* participants with 13 or more ideas and 11 *below average* participants with 12 or less ideas. Results do not reveal any significant differences neither of *above average* participants between the wall condition (*M* = 15.46, *SE* = .45) and the table condition (*M* = 15.77, *SE* = .83), nor of *below average* participants between the wall condition (*M* = 9.09, *SE* = .5) and the table condition (*M* = 9.73, *SE* = 1.45).

Interruptions, Judgments and Deviation from the Topic As in the first study, individual and group warnings had to be used rarely. Individual warnings were displayed seven times, thereof five times in one group. All of these warnings were shown in the table condition. Group warnings were displayed three times, twice in the table and once in the wall condition.

Preferences When asked, which condition participants preferred, 54% stated the table condition and 42% the wall condition (one participant did not have any preference). Reasons that were invoked for the table were that the feedback was better visible (29%), that it was easier to make and hold eye contact (18%), that the group mirror was better integrated in the brainstorming process (13%) and that the feedback on the table enabled better collaboration and communication (13%). One participant stated in the interviews: *"It obviously facilitated face-to-face communication, as you have the feedback system and the other group members in your range of vision at the same time. Thus, it was easier for me to hold eye-contact."* (G1, P1). Reasons stated for choosing the wall condition were that is was perceived as less distracting (25%). Those who sat directly in front of the wall praised the good visibility (8%). Other reasons were that the feedback on the wall produced less pressure than the feedback on the table as it can more easily be ignored (4%). One participant remarked: *"There is less pressure on you through the wall feedback. You are more aware of the feedback on the table and a competition arises to overtake the others (...)"* (G3, P3).

Seating Position The different seating positions, facing the wall or sitting in a 90 degree angle to the wall, did not expose great differences. The number of ideas, for instance, did not differ significantly. To evaluate this, we compared the person sitting on the left (*M* = 12.63, *SE* = 1.34) with the person on the right (*M* = 12.5, *SE* = 1.39). As we did not find any significant differences, we compared the average number of ideas of these participants (*M* = 12.56, *SE* = 1.33) with the average amount of ideas of the participants seated in the middle and thereof facing the wall (*M* = 12.75, *SE* = 1.21). Again, we could not find significant differences. However, when evaluating the qualitative results, one difference was revealed. People facing the wall emphasized that they *"had a better view"* (i.a. G8, P2) and could *"more easily take an occasional peek"* (G3, P2) on the group mirror.

Summary and Discussion

The second study on GROUPGARDEN compared the initial wall version with a tabletop version, with the goal to understand the different affordances and effects of these display environments on brainstorming.

The quantitative measurable data did not reveal significant differences between the two versions. The number of ideas and the balance of participation seemed to be fairly equal in both conditions. Other rules of brainstorming, not to interrupt each other, not to judge ideas of others and to stay on topic, were observed well in both conditions.

The qualitative results however revealed different advantages of both versions. The wall version was perceived as less disrupting and to produce less pressure. The table version in contrast facilitated eye contact and seemed to be a better suited environment for giving the group a sense of easy communication and collaboration.

Overall, these two studies show that group mirrors can effectively support a creative task such as brainstorming. The placement of the visual feedback can have different effects on how group members perceive collaboration. Dependent on the situation, a wall or a table display may be better suited. For instance, for tasks that need group members to work very focused and concentrated, a wall display could be the better choice as it is perceived as less disrupting. The results, furthermore, indicate that different design decisions can make a difference on how group mirrors are accepted. Due to that, we investigate below the aspects of different concepts of visualizations (more cooperative or more competitive) and another aspect of the display environment (private and shared displays).

5.2 A Comparison of Cooperative and Competitive Visualizations

After having investigated the display environment, I will now turn to a study that focuses more on the concepts that underlie the visualization of group mirrors. In this section, I will therefore describe the concept and design of a prototype of a group mirror that comes in three different versions: using a *cooperative*, a *competitive* and a *mixed* visualization (the latter is a mix of the *cooperative* and the *competitive* visualization). Afterwards, I will present results of a study comparing these three versions to each other and to a baseline.

5.2.1 Background and Motivation

When conducting the study on GROUPGARDEN, we observed that several of the participants mentioned to feel under pressure, that they perceived a pressure to succeed and that sometimes a competitive behavior arouse. In related research, these effects of group mirrors have also been mentioned (e.g. Bachour et al., 2008; Schiavo et al., 2014). We assume that this can mainly be attributed to the fact that participants are able compare their own performance with the performance of the other group members. This suggests that the performance increase comes at the cost of an increased feeling of being under pressure and competition between the group members. Thus, we are interested, if a pure group representation (such as the tree in GROUPGARDEN) or a mix with individual and aggregated representations (such as the combination of the tree and the flowers in GROUPGARDEN) can have a similar effect on performance without or with less of the negative effects observed in the previous studies.

Hereafter, I will provide a short classification of existing group mirrors regarding the concepts they use. Before that, I want to clarify the terminology that I will use throughout this section. I will call systems that allow group members to compare their own performance to the performance of the other group members (such as with the flowers of GROUPGRADEN) *competitive*. Group mirrors that only display a group representation (comparable to the tree in GROUPGARDEN) will be called *cooperative*. Systems that combine both approaches are categorized as *mixed* forms (thus, the whole visualization of GROUPGARDEN, with the combination of flowers and tree, is a *mixed* visualization).

In the design space, these factors are not listed as a separate category. They are combinations of different aspects of the design space: The *competitive* visualization is a combination of *individual* and *identifiable* feedback, the *cooperative* version a combination of *anonymous* and *aggregated* feedback.

After discussing the concepts of previous group mirrors, I will illustrate the importance of cooperative behavior in group work by discussion of related research from other domains such as psychology, sociology and economics.

The Competitive Character of Group Mirrors

As described in Chapter 4, most existing group mirrors follow the approach of displaying *competitive* feedback. The system by Sturm et al. (2005) does this by displaying information about speaking times and gaze behavior in front of the participants on a table. Group members can compare their performance to the performance of the others by looking at the visualizations displayed in front of the other group members. The CONVERSATION CLOCK (Bergstrom and Karahalios, 2007a) and CONVERSATION VOTES (Bergstrom and Karahalios, 2007b) represent speaking times of group members with different colors. By looking at the visualization, group members are able to estimate how much they participated in comparison to the others. The MEETING MEDIATOR (Kim et al., 2008) also represents participants with different colors, the balance of participation can also be read from the displays and makes a comparison possible. Streng et al. (2009) uses metaphors to represent group members. The group members that are represented by trees can directly compare their performance, the representation by the weather makes an indirect comparison possible as the mapping is straightforward (bad weather = leafless tree, good weather = flourishing tree). The group mirror of Schiavo et al. (2014) shows the same information to everyone (i.e., who the least attended group member is), though using slightly different kinds of visualization for each group member. Finally, the only group mirror that includes a version that represents the performance of the whole group is the FAN version of the SECOND MESSENGER (DiMicco et al., 2006) (see Figure 2.3). This visualization summarizes speaking times of all group members and shows the deviation from a perfectly balanced discussion. However, this visualization was not understood very well by the participants, thus, it is difficult to make estimations about the influence of this version on group processes.

Motivation, Cooperation and Competition

Cooperation and collaboration are essential elements of human behavior. Barnard (1938) states that the *"origin of cooperation comes from the biological limitations of humans that* *can most effectively be overcome by cooperation."* There are several definitions of the terms of cooperation and competition, for instance, by Maller (1972) or May and Doob (1937). Mead (1937) defines: "*Competition: the act of seeking or endeavoring to gain what another is endeavoring to gain at the same time. Cooperation: the act of working together to one end.*" Another important definition originates from the social interdependence theory of Deutsch (1949). He describes cooperative situations when there is positive interdependence, meaning a positive correlation between the probability of an individual's goal attainment with the probability of another person's goal attainment. Consequently, negative interdependence means a negative correlation.

There is evidence that competitive behavior can have benefits compared to cooperation and that it can be conducive in certain situations (see e.g. Johnson et al., 1978; Sherif, 1976). However, decades of research have also demonstrated the manifold values and advantages of cooperation. Overviews over the literature can be found in Maller (1972), Deutsch (1949), May and Doob (1937) and Johnson and Johnson (2005, 2011). Deutsch (1949) conducted a study in which two groups, either in a cooperative or competitive condition, had to perform a problem solving task. In the cooperative condition, groups were rated against other groups, in the competitive condition, group members were rated against each other. He could show an increase of productivity for the cooperative condition. Johnson and Johnson (1989) could show that cooperation leads to higher achievements and greater retention. Cooperation can, moreover, have a positive influence on higher-level reasoning, on the attitude towards a task and the willingness to take difficult tasks.

Another closely related research area is specifically dedicated to understanding cooperative and competitive reward structures, which is done from a variety of different perspectives, for instance, from the standpoint of psychology, sociology, economics or anthropology. The influence of these reward structures has been investigated in the context of known social processes such as social loafing or social facilitation (Diehl and Stroebe, 1987; Geen, 1991; Karau and Williams, 1993). Especially important for our study is the work of Beersma et al. (2003), who revealed potential benefits of both reward structures. Cooperative rewards (in their case a monetary reward for the whole group) were beneficial for the accuracy of the task (a computer simulated military task) while competitive rewards (monetary rewards only for the best performing individuals) was beneficial for speed. More interestingly, a group with extrovert and agreeable group members made most profit from the cooperative rewards while groups with less agreeable and more introvert group members performed better with competitive rewards.

The approach of group mirrors differs from reward structures as group mirror visualizations do not promise any kind of reward (such as monetary rewards) to the group members. In contrast, group mirrors should serve as subtle support for the group - that can even be ignored when the group wishes. However, it is an interesting question, how much group mirrors actually have in common with or differ from (external) rewards and incentives. Thus, I will briefly summarize recent developments in research on motivation, incentives and rewards. Gerhart and Fang (2015) summarize the development from theories that had a primarily critical view on external rewards to theories that also recognize the potential positive effects of such rewards.

Deci and Ryan (1975) distinguish between *external* and *internal motivation* in their Cognitive Evaluation Theory (CET). *"The term* extrinsic motivation *refers to the performance of an activity in order to attain some separable outcome and, thus, contrasts with* intrinsic motivation*, which refers to doing an activity for the inherent satisfaction of the activity itself."* (Ryan and Deci, 2000). Traditionally, external rewards have been seen as detrimental to *intrinsic motivation* (which has been seen to be of higher quality than *extrinsic motivation* (Gerhart and Fang, 2015). Early work in the field of creativity assessed the influence of external rewards similarly (see e.g. Amabile, 1983, 1996). Amabile (1983) states that *"a primarily intrinsic motivation to engage in an activity will enhance creativity, and a primarily extrinsic motivation will undermine it."*

However, beginning with the Self-Determination Theory (SDT) (Ryan and Deci, 2000), the view on external rewards such as PFP (Pay For Performance) and *extrinsic motivation* has changed. Different forms of *extrinsic motivation* can have different effects. Integrated and identified external motivation as well as internal motivation are categorized under the term *autonomous motivation* while extrinsic motivation in form of external and introjected motivation are labeled as *controlled motivation*, which is seen as lower in quality compared to autonomous motivation. As already mentioned before, more recent work provides evidence that rewards can have a positive impact on intrinsic motivation in some cases (e.g. when it is used in an autonomy-supportive climate) (Cerasoli et al., 2014; Gagné and Deci, 2005). The same change can be observed in research on creativity. Amabile (1993) was able to show that rewards can increase creativity and enhance intrinsic motivation, especially when people were already intrinsically motivated.

Still, the question remains how group mirrors affect motivation in comparison to rewards. While rewards, such as PFP, are normally perceived as positive when achieved, group mirrors do not necessarily reward the group members. Especially when mirroring performance-related aspects of individual participants, some group members might feel rewarded by the group mirrors or, on the contrary, they might even perceive the visualization as a social punishment. Furthermore, the group mirror visualizations are designed to stay in the periphery of the attention so that they can even be ignored - which is less likely for rewards that are promised.

Despite the unknown connection between group mirrors and rewards, it might be valuable to impose similar standards to group mirrors as to rewards. Following the suggestions of Amabile (1993) and Hennessey and Amabile (2010) this would mean that group mirror visualizations should provide *"useful information in a supportive way"*. However, this assumption still needs further evaluation.

Summarizing, we made the observation that most group mirrors use *competitve* visualizations. However, research from a variety of different fields point in the direction that cooperation has a lot of advantages in comparison to competition. Research on reward structures has shown benefits for both types of rewards. However, mixed forms of rewards have, as

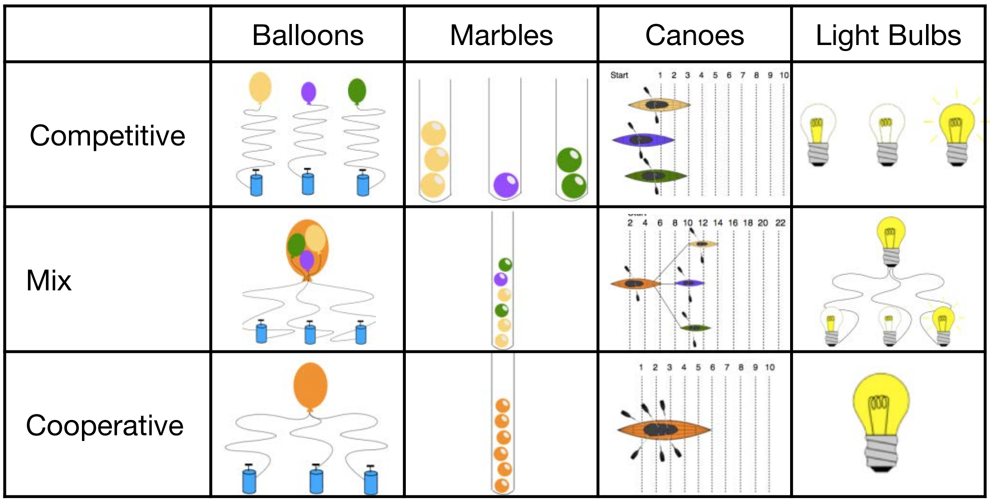


Figure 5.10: Visualizations of the prestudy. Four different visualizations, each in the three versions *competitive*, *mix* and *cooperative*, were evaluated in a prestudy.

far as we know, not been investigated yet. With the study described in the following, we wanted to investigate, which effects other forms of group mirrors, namely *cooperative* or *mixed* visualizations have on groups.

5.2.2 Concept and Design

In a first step, we explored different metaphorical visualizations to find a suited one for the purpose of this study. This was done with a short prestudy. Based on that, a concept and design was chosen, discussed with a group of experts and after another iteration it was implemented.

Initial Design Decision & Prestudy

We decided to position the group mirror on a wall, as this was perceived as less disrupting compared to a tabletop version, even though both display settings would have been suitable, as the study with GROUPGARDEN showed. Based on the results of the study by Streng et al. (2009), we decided to use a metaphoric visualization instead of a diagrammatic one. Our main goal when designing the tool was to find a concept and corresponding visualization that is easy to understand and can effectively represent the different characteristics (*competitive*, *cooperative* and a *mix* between that two). We created four different visualizations (see Figure 5.10), each in three different versions.

The balloon visualization ("change of size") represents individual group members in the *competitive* version with three differently colored balloons that grow. In the *mixed* version, the three balloons are inside of a larger balloon that represents the group's performance. In the *cooperative* version, only the large balloon is displayed. With the marbles, the concept of a “change of amount” is realized. Group members collect marbles, in the *competitive* version colored differently and stored in different containers, in the *mixed* version stored in one joint container. In the *cooperative* version, all marbles have the same color. The canoes realize the concept of “change of distance”, in the *competitive* version, group members are represented with individual canoes, in the *mixed* version, the canoes drag a joint canoe and in the *cooperative* mode, all group members “sit” in one canoe. The distance is indicated with numbers at the top of the visualization. Finally, the *light bulbs* realize the concept of a “change of color”. They light up step by step, either for each participant individually, or connected to a joint light bulb that lights up dependent on the individual ones, or with only one light bulb for the whole group.

We evaluated the four visualizations in a prestudy with 4 participants (1 female, average age 44, 2 engineers, 1 student, 1 teacher). Participants looked at the printouts of the visualizations one after another. The order was counterbalanced using a Latin square design. In short interviews, we asked participants, how they interpreted the visualization in a brainstorming context. Besides that, they filled out questionnaires in which they should, for example, state how easy they can read their amount of ideas of a brainstorming session, if they perceive the success of others also as their own or how strong they perceive a sense of community. We categorized these questions in two categories: (1) supporting a “sense of cooperation” and (2) supporting a “sense of competition”. We summarized the values for each of these categories and built the mean to estimate, which of the visualizations supports these characteristics best. For example, the visualization that supported the “sense of cooperation” best was the *cooperative* version of the marbles ($M = 4.81$). A “sense of competition” was perceived to be best supported in the *competitive* version of the balloon visualization ($M = 4.54$). We calculated scores for all visualizations regarding these characteristics. Results indicate that the balloon visualization represented the concepts of competition and cooperation best. In the interviews, it was furthermore revealed that the balloon visualization was easy to understand. Canoe and marbles were rated second and third best for our purpose.

We presented the balloon visualization to a group of experts (1 female, 6 male; 1 PostDoc, 4 PhD students, 2 professors of HCI, one of them with significant knowledge in Information Visualization). They liked the general concept but commented that the metaphor of balloons inside of balloons is not very realistic. In all three versions, small changes might stay unnoticed due to change blindness[3].

Final Design

The balloon visualization was then adapted according to the feedback of the experts (see Figure 5.11). Additionally to inflating the balloons, small dots indicate the number of ideas of a brainstorming session to reduce change blindness effects. In the *mixed* version, individual participants are now represented with differently colored areas, resembling differently colored gases, instead of individual balloons.

[3] *“The inability to detect changes to an object or scene.”* (Simons and Levin, 1997)

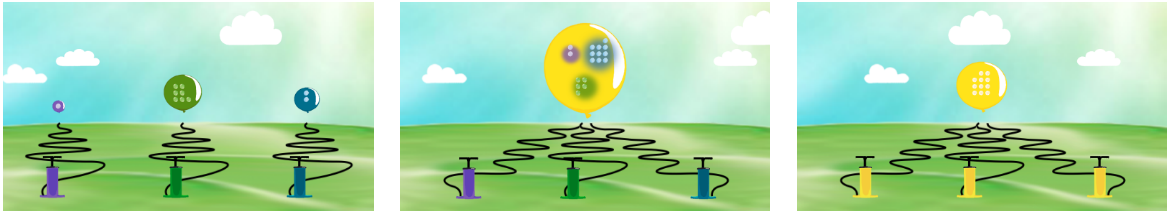


Figure 5.11: Final visualization. The final visualization in its three versions: The *competitive* version (left), the *mixed* version (middle) and the *cooperative* version (right).

In the *competitive* version, all group members are represented with one distinct air pump and balloon, colored differently to make the assignment easier. In the *mixed* version, differently colored gases represent individual users inside of one big balloon that shows the performance of the whole group. Here, the air pumps all lead to the big balloon. In the *cooperative* visualization one big balloon represents the group and the air pumps are all colored equally. When a new idea is stated, a short animation is shown: the air pump moves, the balloon grows and a new small dot is added. In the *cooperative* version, the air pump that moves is chosen randomly.

The system was implemented using Objective C and XCode[4] for iPhones. In the study, the system ran on an iPhone 5c with the operating system iOS7. The experimenter served as a Wizard of Oz and controlled the group mirror. To keep the interaction for the experimenter as simple as possible, a tap on the corresponding air pump triggered the apposition of a novel idea. As the interaction was realized with a simple touch gesture, the same version the experimenter saw was also projected on the wall. We therefore connected the iPhone to a projector that then projected the visualization on the wall.

5.2.3 Evaluation

We conducted a laboratory study to compare the three versions to each other and to a baseline without visualization. Our main questions were, which influence the three different visualizations have on group performance and how well the groups accepts them.

Method

The study was conducted using a repeated measures design with four conditions: three conditions with one of the visualizations each and one baseline condition without any further support. Groups were given four different topics to brainstorm on. Conditions and topics were counterbalanced using a Latin square design, thus each combination occurred once in each round.

[4] https://developer.apple.com/xcode, last accessed: 08.08.2016

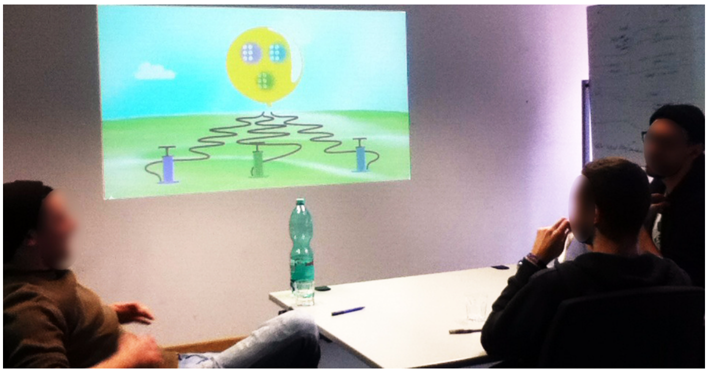


Figure 5.12: Study setup. A group brainstorming with support of the mixed visualization.

Setup and Procedure

The study took part in a quiet room with a rectangular table the three participants sat around. The group mirror was projected on a wall so that all group members faced each other but could still see the visualization (see Figure 5.12).

The experimenter gave a brief introduction to the study and then explained the brainstorming rules. Participants were asked to follow the rules, but they were not explicitly asked to participate in a balanced way. The visualizations were explained before each condition and the topics were introduced to the group. Four topics were chosen: (1) ideas for an app supporting sports and fitness, (2) ideas for an app supporting healthy nutrition, (3) ideas for an app for children and (4) ideas for an app for planning a journey.

During the study, groups brainstormed for eight minutes in each condition. Brainstorming sessions were conducted verbally, without using sticky notes or taking handwritten notes. The study was run as a Wizard of Oz experiment (Kelley, 1983), as already explained in the study with GROUPGARDEN (see Section 5.1.3), thus the experimenter operated the system.

As the experimenter had to assess what counted as an idea in real time, we defined an idea (equally as in the study with GROUPGARDEN) as an on-topic contribution that is novel in the context of this brainstorming session (i.e., it was not stated before). Additionally, ideas building on the ideas of others need to include a somehow novel facet to be counted as an idea. During the first study, two coders rated all sessions in real time. Afterwards, both coders additionally coded the video recordings to count contributions that they did not count as an idea. Cohen's kappa showed substantial agreement ($\kappa = .66$).

After each session, pen-and-paper questionnaires with 5-point Likert scales (1 = strongly disagree, 5 = strongly agree) were handed out. Additionally, we included a questionnaire about six emotions from the PANAS-X questionnaire[5] (negative emotions, positive emotions, guilt, self-assurance, shyness and fear), using the German translation[6].

[5] http://www2.psychology.uiowa.edu/faculty/watson/PANAS-X.pdf, last accessed 28.07.2016

[6] http://www4.ncsu.edu/ dgruehn/page7/page10/files/panas-x-german.pdf, last accessed 28.07.2016

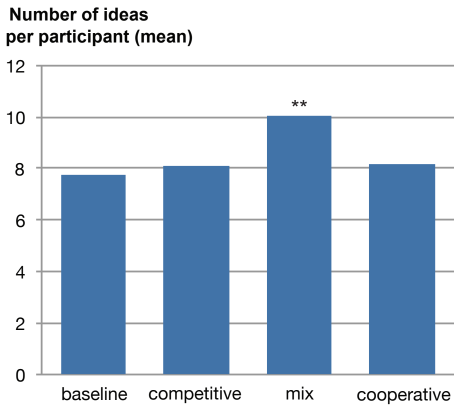


Figure 5.13: Results on the number of ideas. In the *mixed* condition, participants produced most ideas.

	Estimate	95% Confidence Intervals	p-value
Intercept	8.19	[6.38,10.00]	
Baseline	-0.44	[-1.76,0.87]	0.5097
Mix	1.86	[0.54,3.18]	0.0065
Competitive	-0.08	[-1.40,1.23]	0.9015
Topic 2	-0.03	[-0.83,0.78]	0.9463
Topic 3	-0.67	[-1.47,0.14]	0.1078
Topic 4	-0.50	[-0.31,0.31]	0.2267

Table 5.1: Results of the linear mixed model. The cooperative condition and topic 1 serve as reference categories.

In the end, participants filled in a final questionnaire and provided information about their demographics. They were thoroughly debriefed after the study.

Sessions were audio- and video recorded. Videos were taken from the front and the back, so that both the participants and the visualization were visible at the same time. Furthermore, the number of ideas with time-stamp and ID of the participant who stated the idea were logged by the system.

Participants

In the study, 36 voluntary participants (6 female; average age 25, range: 19 to 31 years) took part, 28 were students, 5 software engineers and 2 employees in other professional fields. In all 12 groups, participants knew each other before the study. They could choose between a 10€ voucher from a well-known online web-store or participate in the study as part of an obligation in their study program.

Results

The study was conducted with 12 groups. Four participants built one group (36 participants in total). All these groups took part in each of the three conditions. However, as the 36 participants were nested in groups of four members each, a “standard” evaluation using a repeated measures ANOVA is not possible. Thus, we used a method that takes into account that participants of our study are nested in groups. We used a linear mixed model with condition and topic as fixed effects and groups as a random intercept to evaluate the quantitative data. A dummy coding was used for condition and an effect coding for topic. The model was fitted using the lme function from the package nlme (Pinheiro et al., 2015). The statistical software R was used for the analysis. Qualitative data was gathered from the questionnaires. The PANAS-X questionnaires were also evaluated using a linear mixed model. We are aware that there are ongoing discussions about the evaluation of Likert scales (Carifio and Perla, 2007; Norman, 2010). To our knowledge, data gathered from PANAS-X questionnaires is in

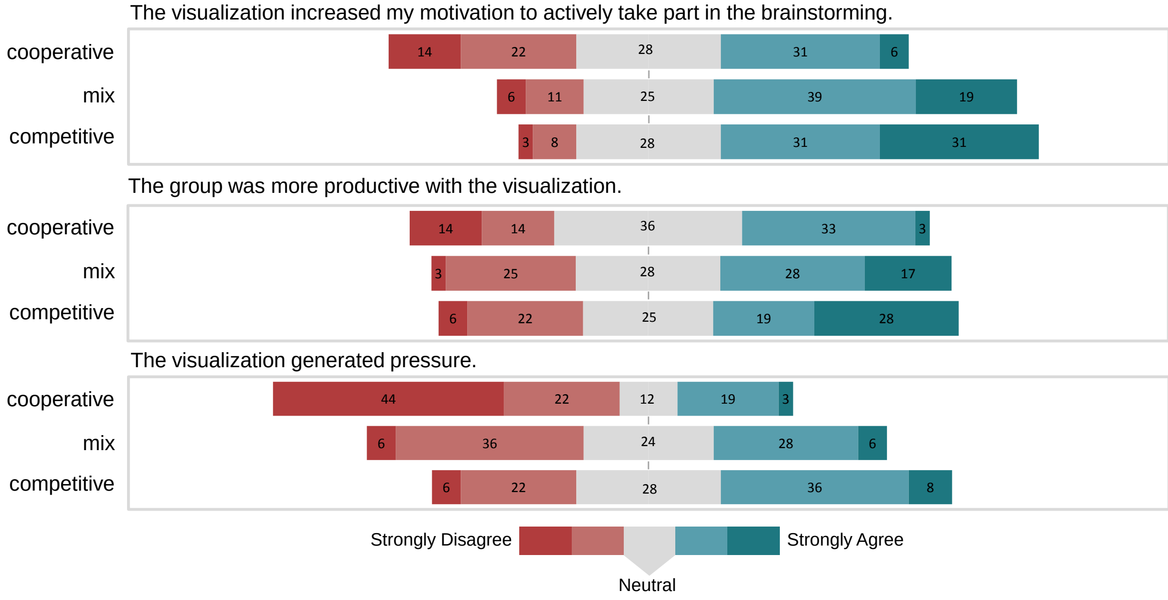


Figure 5.14: Results from the questionnaires. Answers to the questionnaires about motivation, productivity and pressure. Numbers indicate the percentage of participants who answered with that score on the 5-Point Likert scale. Numbers are rounded and thus might not add up to 100% exactly.

related research mostly interpreted as interval data. By reasons of consistency and comparability, we decided to analyze results from the PANAS-X using a linear mixed model.

Performance Calculating the linear mixed model revealed a significant difference between the mixed conditions and the other three conditions. The average number of ideas per participant amounted to 7.75 in the baseline, 8.19 in the cooperative, 10.05 in the mixed and 8.11 in the competitive condition (see Figure 5.13 and Table 5.1). We compared the residual variance between our final model and a model, which only included the group effect as random intercept without any covariables as a measure of the goodness of the model. The residual variance in the full model is 8.13 (variance of random intercept: 7.46), whereas the variance in the null model is 9.22 (7.27).

Balance of Participation To estimate the balance of number of ideas, we used a variation of the Gini coefficient (Weisband et al., 1995), following the example of other researchers (see e.g., DiMicco et al., 2007; Martinez et al., 2011; Schiavo et al., 2016). The equation for a group of three participants is:

Balancing index (BI) = $\frac{3}{4} * \sum_i |participation_i - \frac{100}{3}\%|$

The resulting value reaches from 0 to 1, 0 meaning perfectly balanced and 1 unbalanced. Calculating the BI for our study shows that there is a tendency of more balanced participation in the mixed and competitive condition (BI = 0.14) compared to the baseline and the cooperative condition (BI = 0.16).

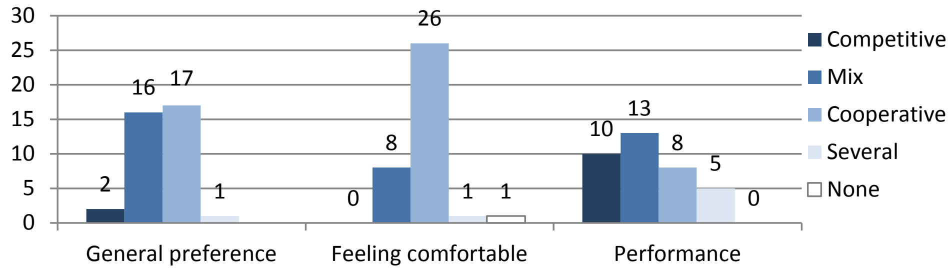


Figure 5.15: Results from the questionnaires. Answers to the questions (1) "Which visualization did you like best?", (2) "With which visualization did you feel most comfortable?" and (3) "With which visualization were you most satisfied with your performance?"

Motivation, Productivity & Pressure In the questionnaires, participants were asked how motivated and productive they felt with the different visualizations (see Figure 5.14). When asked, if the visualization increased the motivation to actively take part in the brainstorming, 22 participants (62%) agreed or strongly agreed for the *competitive* visualization, 21 (58%) for the *mixed* and only 13 (37%) for the *cooperative* visualization. When asked, if participants perceived their group more productive with visualization, 17 participants (36%) agreed or strongly agreed for the *competitive* version, 16 participants (35%) for the mixed and 13 (37%) for the *cooperative*, however here, only 3 agreed strongly. On the contrary, only 8 participants (22%) agreed or strongly agreed that the *cooperative* visualization creates pressure, while 12 (34%) agreed or strongly agreed that the *mixed* and 16 (44%) that the *competitive* creates pressure.

Preferences In the final questionnaire, participants should choose between the three visualizations regarding different aspects (see Figure 5.15) and should justify their choices. We asked, which visualization they liked best. Both the *cooperative* with 17 votes (47%) and the *mixed* with 16 votes (44%) were in the lead. Reasons that were mentioned in favor of the *cooperative* version were that "*the collective idea generation strengthened the team spirit*", (G1, P2) that a "*joint goal*" means "*enjoying teamwork*" (G5, P2) and that it created "*no pressure, therefore much better working and free thinking*" (G12, P3). The *mixed* visualization was liked because of the combination of individual and group feedback: "*You were integrated in the group but at the same time encouraged to contribute*" (G7, P2).

Secondly, we asked participants in which condition they felt most comfortable. 26 participants (72%) stated that this was the case with the *cooperative* visualization, 8 (22%) voted in favor of the mixed, one mentioned several and one none of the visualizations. The reasons that were brought for the *cooperative* visualization mostly included that it produced *"less pressure"* and leads to a *"relaxed atmosphere"*.

When asked, in which conditions participants were most satisfied with their performance, 13 (36%) answered with the *mixed*, 10 (28%) with the *competitive*, 8 (22%) with the *cooperative* and the rest mentioned several. Most participants gave generalizable answers that showed that they perceived the whole group as more productive. Only one participant answered that it was "*because I had a larger idea-area compared to the others*" (G1, P3),

	Estimate	95% CI	p-value
Intercept	1.28	[1.15,1.40]	
Baseline	-0.11	[-1.22,-0.01]	0.0404
Mix	1.03	[-0.14,0.07]	0.5699
Cooperative	-0.11	[-1.21,0.00]	0.0513
Topic 2	-0.01	[-0.07,0.05]	0.7517
Topic 3	-0.03	[-0.09,0.03]	0.3651
Topic 4	-0.01	[-0.05,0.08]	0.6578

Table 5.2: Results of the linear mixed model for negative emotions. The competitive condition and topic 1 serve as reference categories.

	Estimate	95% CI	p-value
Intercept	1.36	[1.21,1.51]	
Baseline	-0.16	[-1.29,-0.03]	0.02
Mix	0.05	[-0.18,0.08]	0.4746
Cooperative	-0.16	[-0.29,-0.02]	0.0228
Topic 2	-0.03	[-0.11,0.05]	0.4656
Topic 3	-0.05	[-0.13,0.03]	0.2523
Topic 4	0.01	[-0.07,0.10]	0.7228

Table 5.3: Results of the linear mixed model for the emotion fear. The competitive condition and topic 1 serve as reference categories.

which is not necessarily related to the specific condition. Reasons for liking the competitive visualizations were mostly that people liked the competitive character. However, here too, some reasons were not necessarily related to the competitive character but the topic (three participants answered that they like the topic about traveling, one the topic about sports and one the topic about nutrition). Reasons for preferring the *cooperative* condition were that participants perceived their ideas as valuable and liked the appreciation of their ideas.

Emotions For emotions, results did not show significant differences regarding the aspects of positive emotions (Competitive: M = 3.64, Baseline: M = 3.60, Mix: M = 3.75, Cooperative: M = 3.67), guilt (Competitive: M = 1.25, Baseline: M = 1.24, Mix: M = 1.18, Cooperative: M = 1.14), self-assurance (Competitive: M = 3.62, Baseline: M = 3.68, Mix: M = 3.74, Cooperative: M = 3.66) and shyness (Competitive: M = 2.45, Baseline: M = 2.42, Mix: M = 2.48, Cooperative: M = 2.41). However, a significant difference could be found for negative emotions (Competitive: M = 1.28, Baseline: M = 1.17, Mix: M = 1.25, Cooperative: M = 1.17) between the competitive condition and the baseline; the difference between the competitive and the cooperative condition is near to being significant. Results are displayed in Table 5.2. Furthermore, a significant difference could be found for the emotion of fear (Competitive: M = 1.36, Baseline: M = 1.20, Mix: M = 1.31, Cooperative: M = 1.20) between the competitive condition and the cooperative condition and the baseline. Results are displayed in Table 5.3.

In summary, results indicate that stronger negative emotions and more emotions linked with fear were reported in the competitive condition compared to the cooperative condition and the baseline.

Summary and Discussion

One main finding of this study is that the performance was better in the mixed condition compared to the other conditions. One attempt to explain this effect could be to take into account the different personalities of the participants. As described in Section 5.2.1, Beersma et al. (2003) found that more extrovert and agreeable group members profit more from cooperative rewards while more introvert and less agreeable group members benefit from competitive reward structures. In our study, a similar effect could have occurred. In the mixed condition, participants who benefit from cooperative structures can focus on the *cooperative* part of the visualization. At the same time, people who benefit from competitive structures can pay attention to the *competitive* part of the visualization. The tendency that participation is at the same time more balanced in the mixed condition indicates that this performance increase can not only be contributed to individual participants but to the whole group.

The results from the questionnaires indicate that group members perceived their motivation highest with the *competitive* visualization and also felt most productive, followed by the *mixed* condition. The quantitative results however show that performance actually was better in the *mixed* than in the *competitive* visualization. One explanation could be that group members felt more motivated, however in the *competitive* condition participants reported about feeling under pressure the most. Furthermore, stronger negative emotions and fear were reported in the *competitive* condition. This might have reduced the positive effects (e.g., in increase in motivation).

Based on these results, we make the suggestion to include a representation of the performance of the whole group in group mirror visualizations. We could show that this increases performance and that participants prefer this version over a solely *competitive* visualization. Furthermore, results indicate that context and goal matter. In most situations, a *mixed* visualization might be appropriate. However, if more is known about the group, for example, that it is composed of people who profit from competitive structures, a competitive version might be beneficial. In situations where it is important that group members feel comfortable, for example, when group members do not know each other, a cooperative visualization could support the group in a better way. However, these are assumptions that require further research to affirm or refute them.

5.3 A Comparison of Public and Private Displays

In the previous sections, two characteristics of group mirrors have been investigated in more detail: the placement of the feedback on table or wall and different concepts underlying the feedback visualizations (more cooperative or more competitive). Results indicate that the different visualizations that revealed more or less information of the other group members influenced performance and attitude of participants towards the system. In this section, a follow up study is described that investigates a related aspect in more detail: the privacy of the display environment. For this, we compared two versions: a *public* and a *private* display setting. I will in the following describe the motivation, the concept of the prototype and results of a study.

5.3.1 Background and Motivation

Results of the study with GROUPGARDEN (see Section 5.1) revealed that group mirrors can exert pressure on group members. The study described in the previous section showed that the design of the group mirror plays a crucial role in that regard. Using a visualization that focuses more on the cooperative aspects of collaboration increased the well-being of participants. Another aspect that could have an influence on the attitude of participants towards the system and performance could be the privacy that the display environment provides.

Most existing group mirrors use public displays, as already described in section 4. Only two systems make use of private displays. The MEETING MEDIATOR (Kim et al., 2008) is a group mirror displayed on mobile devices. However, it displays the same information on all screens. Schiavo et al. (2014, 2016) also employ private displays. Their system does not show the same visualization to everyone, but the same information is encoded in the visualization (the person that is looked at the least sees another visualization than the other group members, however all group members receive the same information: who the person is that is looked at the least). One system can be classified as semi-public: the system by Sturm et al. (2005) shows all information on a table, however, the personal territories of the group members are used to display the information. Thus, it is possible to see information of other group members but not as easy as when the information is displayed in the middle of the table or on a wall. Summarizing, to our knowledge most group mirrors use public displays, few use private displays and these are used to show information that is available to everyone while private displays that only reveal certain information to certain group members have not been investigated in the context of co-located collaboration.

5.3.2 Concept and Design

When distinguishing *public* and *private* display environments for group mirrors, different possibilities exist. The characteristics of *private* and *public* display environments take full effect when combining them with *individual* and *aggregated* representations of the feedback. Figure 5.16 shows the different combinations of *public* and *private* displays and *individual* and *aggregated* feedback. The schematic depiction shows a table on which feedback for a group of three participants is displayed. The smaller rectangles represent *private* displays (e.g., tablets).

The upper row shows how feedback can be represented on a *shared* display. The first depiction shows only *individual* feedback. In this example of a brainstorming session of three participants, the first person (from left to right) and the second person stated three ideas, the third person two ideas. This is the same scenario as the *competitive* version used in the previous section. In the case of a combination of *individual* and *aggregated* feedback, additionally to the *individual* representation, the performance of the whole group is displayed (in this case seven red dots representing the overall amount of ideas of the group). This is equal to the *mixed* version used in the previous section. The third possibility of representing

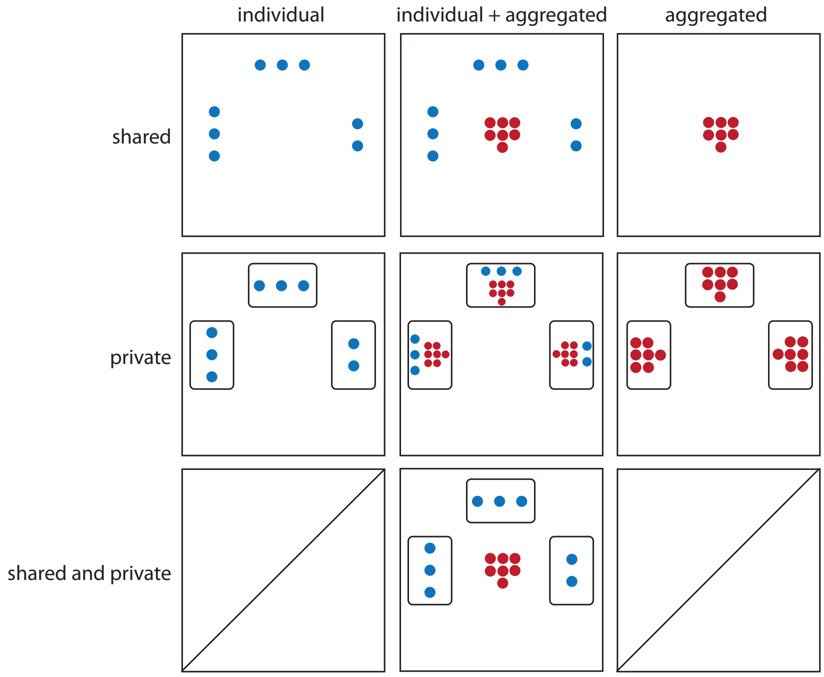


Figure 5.16: Different possibilities of combining shared/private and individual/aggregated feedback. The blue dots represent the number of ideas during a brainstorming session of three participants (i.e., the person on the left and the person in the middle stated three ideas, the person on the right stated two ideas). The red dots represent the number of ideas of the whole group (seven in total).

information on a *shared* display is to do it using only *aggregated* feedback, i.e., only showing the overall amount of ideas. This is the same scenario as the *cooperative* version used in the previous section.

The middle row shows how feedback can be represented on *private* displays. When using only *individual* feedback, each group member can see how many ideas he or she stated but has no visual indication of how many contributions the others made. In case of *individual* feedback in combination with *aggregated* feedback, each group member sees on her or his *private* display how many ideas she or he contributed in relation to the overall number of ideas of the group. The case that *aggregated* feedback is shown on *private* displays is similar to showing *aggregated* feedback on a *shared* display, as the information content is the same.

Finally, the lower row shows how *shared* and *private* displays can be used in combination. Showing only *individual* or only *aggregated* feedback does not require a combination of shared and private displays. However, showing both *individual* and *aggregated* feedback can be done using a combination of *shared* and *private* displays: *individual* feedback can be shown on the *private* displays while *aggregated* feedback is shown on the *shared* display. This contains the same information content as the combination of *private* and *individual* + *aggregated* feedback.

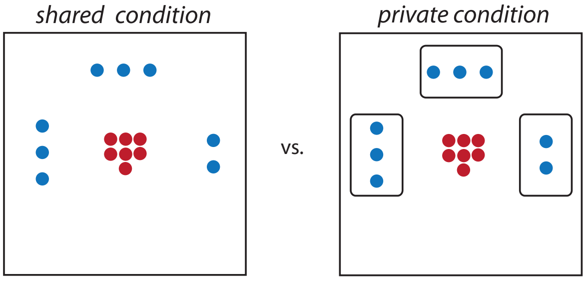


Figure 5.17: Study design. For the study, we chose to compare a *shared condition* and a *private* condition, both showing individual and aggregated feedback. In the *shared* condition, individual and aggregated feedback is displayed on a table. In the *private* condition, individual feedback is shown on private displays while the aggregated feedback is displayed on the table.

Our goal was to understand the differences between a *shared* condition in which all group members can see all information (we therefore chose the *individual + aggregated* version) and a *private* condition in which *individual* information is only shown to the individual participants. As this should be the only dependent variable, we chose the version using *shared* and *private* displays to show *individual* and *aggregated* information. This led us to the comparison shown in Figure 5.17 and 5.18.

To decide, which visualization to choose for the study, we conducted a short focus group study. Four different metaphorical visualizations were discussed during the focus group, one of them the balloon representation used in the study on *cooperative* and *competitive* visualizations. More details can be found in Sachmann (2014). Based on the results of the focus group and for reasons of consistency and comparability with the study described in the previous section, we chose the balloon visualization for the main study.

The system was implemented as a web application using HTML5 and the Javascript library create.js[7] to enable platform independent use. For the study, we used three Samsung Tablets (two Galaxy Tab2, one Galaxy Tab3) with OS Android as private displays and a desktop PC for realizing the shared display. The shared visualization was projected on the table. A client-server architecture was implemented using PHP and AJAX.

5.3.3 Evaluation

The main goal of the study was to investigate the differences between *private* and *shared* displays for group mirrors. The main questions were, which influence more *private* and more *public* display environments have on group performance and how well the group accepts them.

[7] http://createjs.com/, last accessed 28.07.2016

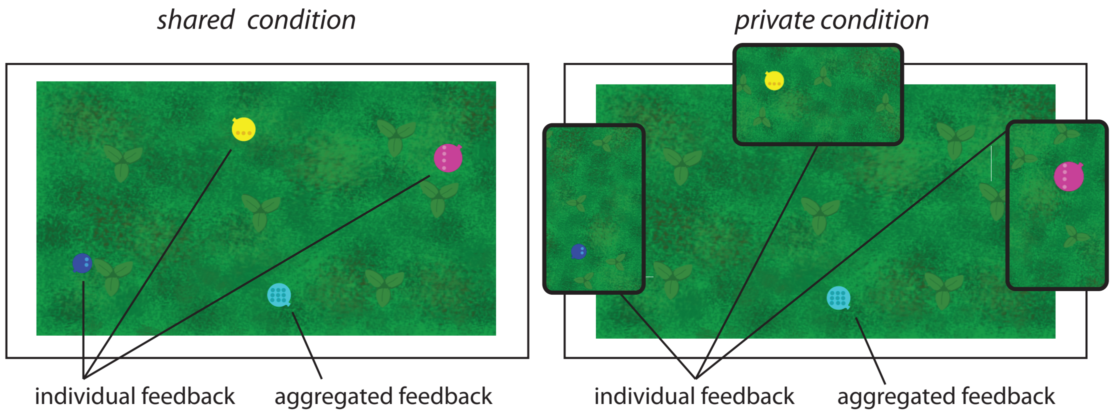


Figure 5.18: Visualizations. The system uses the metaphor of balloons that grow depending on how many ideas are stated. In both conditions, the amount of ideas of individual group members is displayed inside of individual balloons that are either shown publicly on a table or privately on tablets. In both conditions, a fourth balloon represents the overall amount of ideas.

Method

The study was run as a laboratory study using a repeated measures design. Three conditions were compared, a baseline without any support of a visualization, a *private* condition in which a part of the visualization was displayed on private devices (tablets) and a *shared* condition in which the whole visualization was displayed on a shared tabletop display. Three topics were used. We counterbalanced topics and conditions using a Latin square design. For a perfect counterbalancing, nine groups would have been necessary. However, for our study, we were able to recruit subjects for six groups only, which should still be enough to achieve relevant results.

Setup and Procedure

The study was conducted in a quiet room equipped with a rectangular table. The group mirror visualization was projected on the table. In the *private* condition, tablets were handed out to the participants (see Figure 5.19).

First, the experimenter gave a general introduction to the study procedure. Additionally, the brainstorming rules were explained. As in the previous studies, participants were asked to try to follow these rules but were not explicitly asked to strive for balancing their amount of contributions. The visualizations and topics were explained before each condition. We used the same topics as in the study described in the previous section (minus one topic, as this study only compared three conditions instead of four): (1) ideas for an app supporting sports and fitness, (2) ideas for an app supporting healthy nutrition and (3) ideas for an app for planning a journey.

Each group brainstormed for ten minutes per condition. Brainstorming sessions were held verbally without taking notes or using sticky notes. The experimenter took the role of a Wizard of Oz (Kelley, 1983) and controlled the visualization. Every time someone states a

Figure 5.19: Study setup. Left: Public condition. Right: Private condition.

new idea, the experimenter increases the idea counter of that person and the visualization is updated, i.e., the corresponding balloons grow.

Therefore, we needed to define an "idea". Equally to the previous studies we defined it as an on-topic contribution that is novel in the context of this brainstorming session (i.e., was not stated before). Additionally, ideas building on the ideas of others need to include a somehow novel facet to be counted as an idea. Two coders coded the first study in real time to identify ideas. Afterwards, both coders also coded the video recordings to identify contributions that they did not count as an idea. In that process, the definition was slightly refined (i.e., single keywords without any explanation are not counted as an idea). Cohen's kappa showed moderate agreement ($\kappa = .42$).

Questionnaires with 5-point Likert scales (1 = strongly disagree, 5 = strongly agree) were handed out after each brainstorming session, including also the questions of the PANAS-X questionnaire[8] about fear, attentiveness, shyness and serenity. We used the German translation of the PANAS-X[9]. After all brainstorming sessions, a final questionnaire was handed out, including questions about demographics.

Sessions were audio and video recorded. Two cameras were positioned in front of the participants, so that all participants and the table were visible on the recordings. The ID of the participants together with the time-stamp when an idea was stated were logged.

Participants

In total, 6 groups took part in the experiment with each 3 participants (9 female; average age 22, range: 19 to 32), resulting in 18 participants in total. 16 were students, 2 employees. All group members of a group already knew each other before the study. Participants were rewarded with a 10€ voucher for an online store. Alternatively, some participated as part of an obligation in their study program.

[8] http://www2.psychology.uiowa.edu/faculty/watson/PANAS-X.pdf

[9] http://www4.ncsu.edu/ dgruehn/page7/page10/files/panas-x-german.pdf

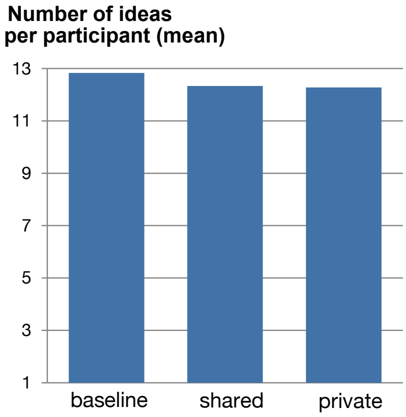


Figure 5.20: Results on the number of ideas. They do not reveal significant differences between the three conditions.

	Estimate	95% Confidence Intervals	p-value
Intercept	12.33	[9.43, 15.24]	
Baseline	0.5	[-2.79, 3.79]	0.77
Private	-0.06	[-3.35, 3.24]	0.97
Topic 1	-0.43	[-2.33, 1.47]	0.66
Topic 2	0.24	[-1.66, 2.14]	0.81

Table 5.4: Results of the linear mixed model. The shared condition and topic 3 serve as reference categories.

Results

The study was conducted with 6 groups with each 3 participants that built one group (18 participants in total). These groups took all part in each of the three conditions. However, as the 18 participants were nested in groups of four, a "standard" evaluation using a repeated measures ANOVA is not possible. Consequently, we used a method that takes into account that participants of our study are nested in groups. We calculated a linear mixed model with condition and topic as fixed effects and group as a random intercept. We used a dummy coding for condition and effect coding for topic. The model was fitted using the lme function from the package nlme (Pinheiro et al., 2015). We used the statistical software R.

Performance The linear mixed model did not reveal any significant differences between the three conditions. The average number of ideas in the baseline was 12.83, in the *shared* condition 12.33 and in the *private* condition 12.27 (see Figure 5.20 and Table 5.4). We compared the residual variance between our final model and a model, which only included the group effect as random intercept without any covariables as a measure of the goodness of the model. The residual variance in the full model is 8.13 (variance of random intercept: 7.46), whereas the variance in the null model is 9.22 (7.27).

Balance As in the study explained in the previous section, we used the balancing index (BI), a variation of the Gini coefficient (Weisband et al., 1995), for calculating the balance of participation. More details can be found in Section 5.2.3. The results indicate that participation was more balanced in the *private* condition (BI = 0.16) compared to the *shared* condition (BI = 0.18) and to the baseline (BI = 0.24).

Motivation, Productivity & Pressure Additionally to the performance related findings we wanted to understand, how participants assessed themselves in the different conditions. In the questionnaires that were handed out after each condition, we asked how pressured people felt to generate new ideas. The results from the Likert scales are depicted in Fig-

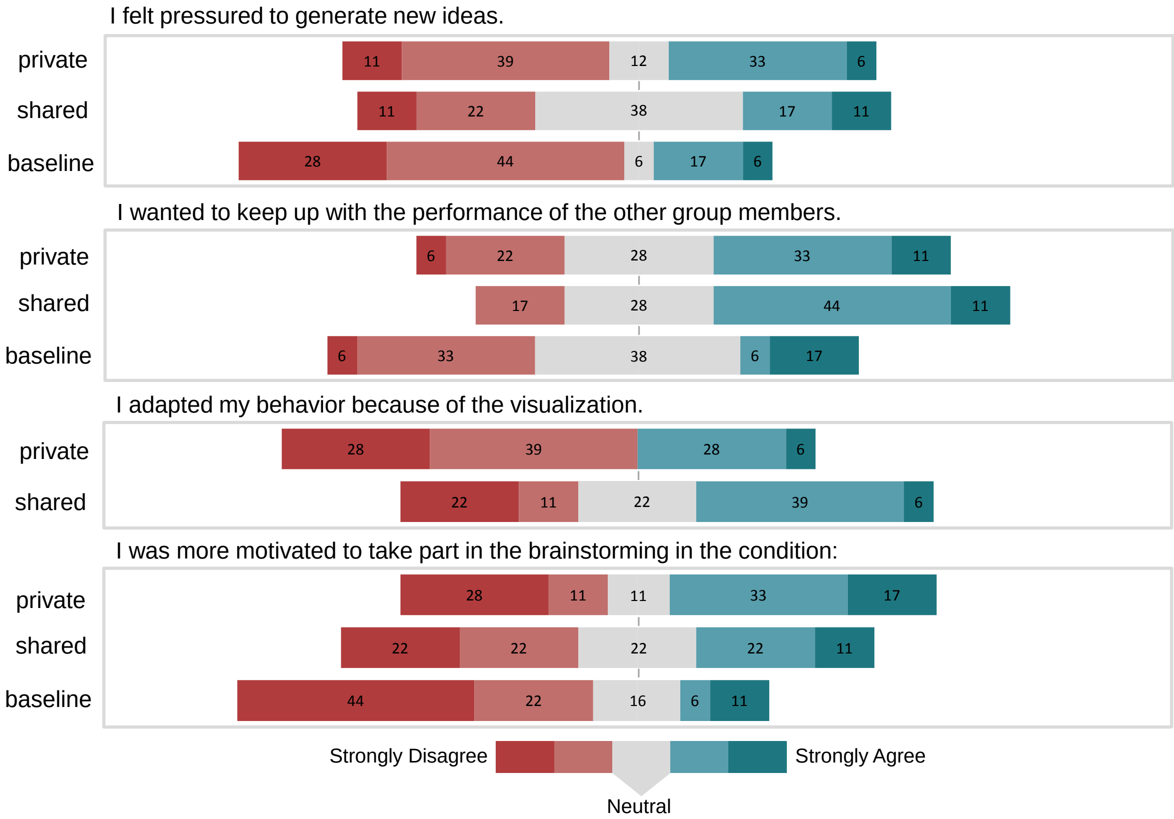


Figure 5.21: Results from the questionnaires. Answers to the questions about pressure, performance and the influence of the visualization on the behavior. Numbers indicate the percentage of participants who answered with that score on the 5-Point Likert scale. Numbers are rounded and thus might not add up to 100% exactly.

ure 5.21. In the private and the shared condition, there were both participants that agreed or strongly agreed that they felt pressured (public: 7 participants (39%), shared: 5 p. (28%)) as well as participants that disagreed or strongly disagreed (public: 9 p. (50%), shared: 6 p. (33%)), only a tendency to more people disagreeing in both conditions is observable. In contrast, in the baseline, more participants disagreed or strongly disagreed (13 p., 72%) to feel under pressure to generate new ideas.

Additionally, we asked participants, if they felt the urge to keep up with the performance of the other group members. Here, a slight difference between the private and the shared condition is visible. While in the private condition only 8 participants (44%) agreed or strongly agreed, it were 10 participants (55%) in the shared condition. Only 4 participants (23%) agreed or strongly agreed in the baseline, when asked if they wanted to keep up with the performance of the others. After both conditions with group mirror, we posed the question if participants adapted their behavior due to the presence of the visualization. In the private condition, 6 participants (34%) agreed or strongly agreed, in the shared condition, 8 participants (45%) agreed or strongly agreed.

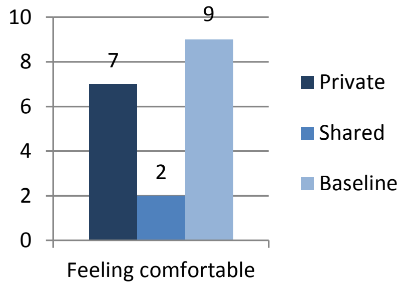


Figure 5.22: Results from the questionnaires. Answers to the question "In which condition did you feel most comfortable?"

In the final questionnaire, we asked participants in which of the conditions they felt more motivated to take part in the brainstorming. Results indicate that people felt more motivated with private displays than in the other conditions (9 p. (50%) agreed or strongly agreed), followed by the shared condition (6 p. (33%)). In the baseline, only 3 participants (17%) agreed or strongly agreed to feel more motivated than in the other conditions.

Preferences We asked participants, which of the display settings they felt most comfortable with. 7 participants (39%) stated that this was the case with the private displays, 2 (11%) with the shared display and 8 (45%) in the baseline condition (see Figure 5.22). Reasons that were brought for choosing the private environment were that *"you have an overview of the amount of your ideas (which is superior to the baseline condition) but at the same time there is less pressure compared to the shared system"* (G4, P3). One reason for the shared display environment was that this person thought that *"the system clearly motivates; with the public system you can compare your performance and generate a competition. I don't mind if others can see and compare my performance."* (G6, P2). When participants preferred the baseline, reasons were that this situation produces less pressure (G1, P2; G5, P2), feels less competitive (G6, P3) and resembles more a *"normal conversation"* (G2, P1). One participant did not recognize any difference to the conditions with group mirror and therefore preferred the baseline, as *"electricity is expensive"* (G3, P1). One participant perceived participation more balanced in the baseline (G5, P3), though this was actually not the case.

Emotions With the PANAS-X questionnaire, we collected information about fear, attentiveness, shyness and serenity. Results of a linear mixed model did not show significant differences between the conditions. In general, values for fear (Shared: M = 1.55, Private: M = 1.28 , Baseline: M = 1.44) and shyness (Shared: M = 1.35, Private: M = 1.50, Baseline: M = 1.42) are similarly low, values for attentiveness (Shared: M = 3.83, Private, M = 3.88, Baseline: M = 3.81) and serenity (Shared: M = 3.22, Private: M = 3.48, Baseline: M = 3.56) generally higher.

Summary and Discussion

In the described study, two different display environments, one using private displays in combination with a shared display and one using only a shared display, were compared with a baseline without any support through a system. The information that was shown on the displays was a combination of individual and aggregated feedback. In the private condition, the individual feedback was displayed on the private displays. In the study on cooperative and competitive visualizations (see Section 5.2), this combination appeared to be most beneficial for performance and was at the same time well liked by the participants. Therefore, our assumption was that the visualization would have similar positive effects as in the study on cooperative and competitive visualizations. Our main goal was to see, if there are further performance related differences when using private or shared displays and to see, with which display setting people feel more comfortable.

We could not find any significant differences regarding the performance, though the results of the study described in Section 5.2 would have suggested that there could at least be differences between the public condition and the baseline. There might be several reasons for this. On the one hand, due to time constraints, the study was only conducted with six groups, which might have been too few to reveal significant differences. On the other hand, we used a slightly different visualization than in the study described in the previous section. The amount of individual ideas (individual balloons) is displayed in front of each group member and the overall amount of ideas (group balloon) in the middle of the table. In the study on cooperative and competitive visualizations, the individual amount of ideas (individual balloons) was displayed inside of the group balloon. We changed this as we wanted to design the different conditions of the study as similar as possible to control possible confounding variables in this particular study. However, seeing the direct relation of the individual contribution and the contribution of the whole group (individual balloons inside of a group balloon) might be more stimulating for idea generation than aligning these information next to each other without showing the connection, as it has been done in this study.

Our results indicate that with visualization, participation was more balanced than without. This confirms results of several studies on group mirrors, e.g., the study on GROUPGARDEN (see Section 5.1). Qualitative results also revealed some differences between the conditions. Participants seemed to feel a bit more comfortable with the private displays compared to the shared display, they seemed to feel less under pressure and several participants stated that they felt more motivated with the private displays. As results did not reveal any negative performance related effects of using private displays, this could be an opportunity to design group mirrors in a more agreeable and pleasant way. In situations in which group members who do not mind comparing their performance to the other group members, for example, when group members know each other well and feel comfortable with seeing that information, the extra effort of using several private displays might not be necessary. However, especially in situations in which group members could feel under pressure, using private displays instead of a shared one to display a group mirror could be beneficial.

5.4 Chapter Summary

This chapter explored the influence of several factors of group mirrors on brainstorming. In one project, the general feasibility of a group mirror to support brainstorming was evaluated. In three further projects, the factors of the display environment (table or wall and public or private) and the concept underlying the group mirror visualization (competitive or cooperative) were evaluated in more detail. In the following, the main findings are summarized.

- One main finding of these studies is that group mirrors actually are a suitable way to support co-located brainstorming sessions. Participants of our studies generally liked the support and a number of positive results could be observed. However, the design of the group mirror is crucial for its success.
- Both table and wall displays are suitable options for showing the group mirror visualization, with different advantages and disadvantages. While a group mirror on a table seems to support communication and collaboration better, a wall display seems to be less disrupting and imposed less pressure on the group.

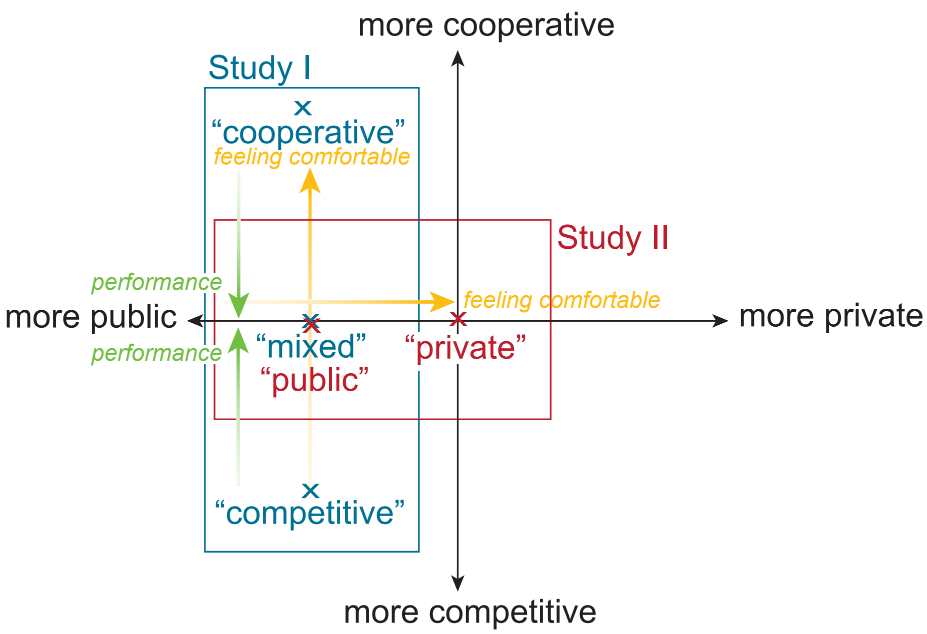


Figure 5.23: Overview of results. Results indicate that people feel more comfortable (yellow arrows) with more cooperative visualizations (study I, in blue) and with more private display environments (study II, in red). Better performance (green arrows) is achieved using mixed rather than only competitive or cooperative visualizations (study I).

- When comparing more competitive visualizations (i.e., visualizations that allow group members to compare their performance with each other), cooperative visualizations (i.e., visualizations that show the overall performance of the group) and mixed forms (i.e., showing both, individual and aggregated information), performance turns out to be supported best with the mixed visualization, while participants feel most comfortable with the cooperative version, followed by the mixed.

- Private displays that show individual information instead of showing this information on one shared display seem to create a more comfortable environment. Using private displays seems to put a bit less pressure on group members while group members are still motivated to participate in the brainstorming.

- Taken together, people feel more comfortable using private displays showing individual information and using more cooperative visualizations. Figure 5.23 schematically depicts the results of two studies. Results on the study on *competitive*, *mixed* and *cooperative* are displayed in blue, results from the study on *shared* and *private* in red. The middle of the coordinate system represents a situation that uses a combination of *cooperative* and *competitive* visualizations and *shared* and *private* displays. The quadrant I (upper right) represents a situation in which a group uses a more cooperative visualization and private displays. In our balloon example this would be the group balloon shown on tablets. This leads to a situation in which even people outside the group such as passers-by cannot see the information. Based on the results of our study, we assume that group mirrors fulfilling these characteristics lead to less pressure and make group members feel more comfortable. This assumptions however still need further evaluation. This will also be discussed in the future work part (see Chapter 9.4.3). in more detail. In terms of performance, results of our study indicate that mixed visualizations are beneficial, differences between public and private displays could not be detected.

6

Supporting the Disney Method

In the previous chapter, the influence of different group mirrors on brainstorming was investigated in detail. In this chapter, another collaborative creativity technique, the *Disney Method* (Dilts, 1995), will be in the focus. This is one of a group of creativity techniques that make use of different roles to stimulate the idea generation process. Letting people slip into different roles can be beneficial for creativity as they do not necessarily need to express their own point of view but can bring up aspects that they normally would not have revealed. Furthermore, using different roles can make groups aware of the various perspectives from which a topic can be approached. The *Six Thinking Hats* Method (De Bono, 1985), for instance, uses differently colored hats that represent the different roles. The red hat is used to express emotions, the green hat stands for creativity, the black hat for discernment. The technique that I will concentrate on this chapter is the *Disney Method*. It uses the roles of the *dreamer*, the *realist* and the *critic*. More details on these roles can be found in Section 3.1.1.

Different methods to reduce the problem of stress and pressure when using group mirrors has been investigated and addressed in the previous chapter. This, for instance, has been done by using a setting with private displays or by using cooperative concepts to visualize information. The group mirrors presented in the current chapter link to this idea of employing more cooperative concepts. Instead of showing individual performance related aspects, the systems mirror more content-related information (i.e., the usage of the different roles). One of the main questions that this chapter tries to answer is, if these types of group mirrors are beneficial for collaborative creativity in terms of performance and acceptance.

This chapter is based on one practical work (Steinberger, 2013) and one bachelor thesis (Nußberger, 2014). Part of it was published in the proceedings of two conferences (Tausch et al., 2015a,b) with the co-authors Fabius Steinberger, Fabian Nußberger and Heinrich Hußmann. Some of the images and tables shown in this chapter have orignially been published in these papers. The detailed personal contribution statement can be found in the disclaimer.

Two other issues are additionally addressed throughout this chapter. The study that is described in the first section was conducted as a field study. The majority of previous studies on group mirrors have been realized as lab experiments. Thus, there is little information on how group mirrors are used in the "real world". This study is an approach to understand the effects of a group mirror "in the wild".

The second issue that we want to address is the missing integration of additional material. For instance, sticky notes are often used to record ideas in brainstorming sessions. This can interfere with the use of a group mirror that is shown on a table or a wall. Thus, in the second section, I will discuss a system that integrates a functionality to enter ideas.

6.1 A Group Mirror for the Disney Method

In this chapter, I will describe a group mirror supporting the Disney Method. After a general introduction in form of the background and motivation to the topic, the concept and design are presented. Afterwards, a field study and its results are outlined.

6.1.1 Background and Motivation

One of the main intentions of this prototype was to investigate another concept of a less competitive and more cooperative visualization including additional information (i.e., the role distribution of the Disney Method's roles). Group mirrors from related research mainly focus on mirroring information that enables group members to compare themselves to the other group members. Several systems allow to see the speaking times of all group members on a public display. However, as discussed in the previous chapters, this can produce pressure for the group. We were interested to investigate another concept that moves away from this direct comparison. We do this by focusing more on content-related aspects. We visualize the amount of ideas that have been stated of a specific role instead of showing information about who stated that idea.

Besides that, we intended to explore this group mirror in a field study instead of a laboratory study. With that, we hope to add to the understanding on how group mirrors could be used for collaborative creativity sessions in real use cases.

6.1.2 Concept and Design

When using the Disney Method for the first time, there are two main challenges. First, group members need to remember the roles and their meanings to effectively integrate them into the creative process. Second, all roles should be made use of. Imbalances of the use of the different roles might occur. For instance, people shy away from the critic role as it means to challenge their colleagues' or friends' concepts and ideas.

Figure 6.1: System design. Left: When a new contribution is made, a small circle with the color of the corresponding role is added to the circles in the middle. A short animation is shown. Right: The visualization is shown on a display that is lying flat on a table.

Information As the key element of the Disney Method is to use the three different roles, the main goal of the group mirror is to encourage groups to use all roles in a fairly balanced way. To support this, the group mirror shows how many contributions to the specific roles have been made. This implies that the overall amount of contributions is visible as well. However, the visualization does not allow to distinguish between the group members.

Visualization Three differently colored circles represent the three different roles. Using circles displayed in the middle of a tabletop display allows every group member a good view on the visualization regardless of seating position or angle and makes it easy to compare the size of the circles from every seating position around the table. The visualization follows an information decoration approach (Eggen and Van Mensvoort, 2009) with the goal to balance aesthetical and informational quality. Thus, the visualization is composed of simple shapes using plain colors. In the beginning, all circles are equally small. They grow the more ideas the group stated for that particular role. A short animation is shown when a new idea is added. A small circle moves from the edge of the display to the middle and joins the corresponding circle (see Figure 6.1, left).

Colors Although the Disney Method in its original form does not assign specific colors to the roles, we decided to use colors to make it easier to recognize the information when displayed in the periphery of the attention. The dreamer is represented with green, the realist with yellow and the critic with red. We labeled the circles with the names of the roles to make it easer for group members to remember the roles and to make the group mirror accessible for color-blind people.

Placement We decided to show the group mirror visualization on a tabletop display. The study described in chapter 5.1 indicated that a group mirror on a table produced less pressure and eased communication. Furthermore, in this particular case, groups were composed of six participants. With a wall display it would be more difficult to place the visualization in a way that it is visible for all participants. Therefore, we displayed the information on a screen that was lying flat on a table. We used a 24 inch display on which the group mirror ran in full screen mode (see Figure 6.1, right).

The group mirror was implemented using Processing[1] to allow platform independent use. The visualization can be controlled via keyboard input. Different keys were assigned to the three Disney roles. When one of these keys is pressed, the circle of the corresponding role grows.

6.1.3 Evaluation in the Wild

We conducted a field study to investigate the effects of the group mirror in a "real world" scenario. Our main questions were, what influence the group mirror has on performance, on the use of the different roles and on the acceptance of the feedback.

Method

As already described above, we conducted a field study, because there are only few studies that investigated group mirrors in the wild. Conducting a field study brings a number of advantages compared to laboratory studies, such as an increased external validity (Rogers, 2011). We compared two conditions, one with support of the group mirror, one without. The study was designed as a within-groups experiment.

Setup and Procedure

We conducted the study during a practical course at our university. The study took part in an intermediate step of this course. Beforehand, students had developed app ideas in groups of four. Now, these ideas should be discussed using the Disney Method. Therefore, each group presented their app idea. After each presentation, the app idea was discussed for ten minutes using the Disney Method. As 16 students took part in the course, four discussions were held in total. The four students that just had held the presentation did not take part in the discussions but listened and took notes. The remaining students were divided in two groups of six participants. One of these groups was supported with the group mirror. In each discussion the combination of participants was altered to reduce the effects of opinion leaders and group dynamics.

As in the studies described earlier, we used a Wizard of Oz approach (Kelley, 1983) to control the group mirror. Two persons, each one sitting next to one group at a time, classified the contributions according their role. All sessions were audio and video recorded. We furthermore handed out questionnaires after each session with 5-point Likert scales (1 = strongly disagree, 5 = strongly agree). Time and role of entered contributions were logged by the application.

[1] https://processing.org, last accessed 06.08.2016

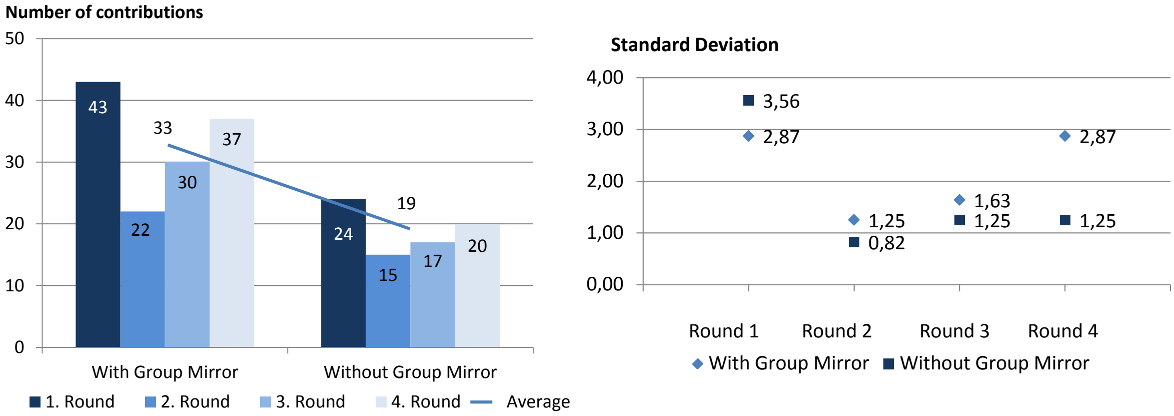


Figure 6.2: Results on the number of contributions and balanced use of the roles. Left: Absolute number and average of contributions with and without group mirror. Right: Standard deviation of the different Disney roles in the different rounds.

Participants

In the study, 16 participants (5 female; age range from 20 to 25 years) took part. All participants were media informatics students participating in a course on iOS development. All participants already knew each other as they at least had been working together on their app idea over the past three weeks. Only one of the participants was familiar with the Disney Method, all others did not know this creativity technique.

Results

The study was evaluated using the questionnaires, audio- and video-recordings and application loggings. A dependent t-test was used to evaluate quantitative information. A 5% level of significance was applied to the tests. Excel was used for calculating the t-test.

Performance With support of the group mirror, more contributions (independent from the role) were stated than without group mirror. We compared the amount of contributions that were voiced with group mirror ($M = 33$, $SD = 9.055$) with the amount of contributions stated without the support of the system ($M = 19$, $SD = 3.916$). A dependent t-test reveals a significant difference between these conditions ($p < 0.05$) (see Figure 6.2, left).

Balanced Use of Roles We could not find significant differences regarding the balanced use of the three roles. To calculate how balanced the roles were used, we compared the distribution of the number of contributions in each role from their mean (i.e., the standard deviation). As Figure 6.2 (right side) shows, the standard deviation was lower without group mirror in three rounds (without statistical significance). However, this can also be attributed to the fact that the number of contributions was higher with group mirror.

When looking at the use of the different roles, it became apparent that our initial assumption that people shy away especially from the critic role (that was based on observations of other groups using role-based creativity techniques) actually was true. Without group mirror, 39%

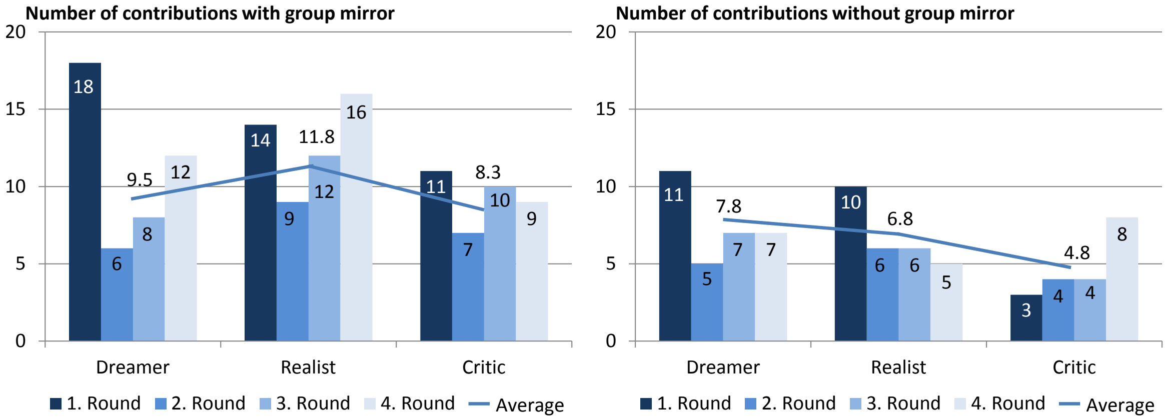


Figure 6.3: Results on absolute number and average of contributions separated between the Disney roles. Left: With group mirror. Right: Without group mirror.

were dreamer contributions, 36% contributions attributed to the role of the realist and 28% to the role of the critic. Furthermore, we could observe a tendency that especially the less used roles of the realist and the critic are used more frequently with support of the group mirror. With system, 33% were dreamer, 39% realist and 28% critic contributions. The absolute numbers are displayed in Figure 6.3.

Social Pressure and Stress As this type of visualization was a first attempt to focus more on content-related aspects instead of performance-related aspects, we were interested, if participants felt pressured through the system. Results from the questionnaires confirm that this group mirror visualization did not put much pressure on group members. Without group mirror, all participants disagreed to have felt stressed during the discussion. With group mirror, 13 disagreed, 2 were neutral, only one agreed that he or she had felt stressed. We furthermore asked, if participants felt observed. In both conditions, participants either disagreed or were neutral regarding this aspect (with group mirror: 11 disagreed, 5 neutral; baseline: 13 disagreed, 3 neutral).

Disruption The group mirror was not perceived as too disrupting. 15 participants disagreed, one was neutral regarding the aspect of disruption. Furthermore, 13 group members agreed that they could follow the discussion and the information on the display simultaneously, 3 were neutral. This is also confirmed by the answers to the question if participants were able to focus on the discussion. In the group mirror condition, 15 participants agreed, 1 ticked neutral. In the baseline condition without feedback, 15 agreed that they could focus on the discussion, 1 participant did not agree.

Preferences In the questionnaires, we asked participants if they liked the condition with or without group mirror more. 14 stated that they liked being supported by the system more than without, one participant liked both conditions equally and one preferred the condition without feedback.

Summary and Discussion

With the prototype and study described in this section, we investigated how a group mirror showing information about the usage of different roles in a role-based creativity technique influences collaboration. We did this by conducting a field study during a course at our university. We compared two conditions, a baseline and a condition in which the group mirror shows, how many ideas to the three different roles, *dreamer*, *realist*, and *critic* were stated.

Results show that the group mirror led to an increase of the overall amount of contributions. In a previous study (see Section 5.2), we compared a visualization that showed the overall amount of ideas (the *cooperative* visualization) with a baseline, however, we could not find significant differences in the amount of ideas. In the study described in the current section, the visualization also shows the overall amount of ideas. Here, the study shows a significant difference in the amount of contributions compared to the baseline. One explanation for these different findings could be that the visualization showing the different roles of the Disney Method has encoded additional information (i.e., the different roles). Being reminded of the different aspects that group members can contribute could have led to the increase in performance, while in the study described in Section 5.2 no additional, possibly stimulating, information was visualized.

However, the study setup could have been another reason for these different findings. In the study described in this section, only two conditions were compared in a within-groups experiment. This implies that half of the participants took part in the baseline condition without knowing a group mirror system before. In the study described in Section 5.2, four conditions were compared in a within-groups experiment. Consequently, only one out of four participants took part in the baseline without knowing a group mirror. Having used a group mirror before could lead to an adapted behavior (i.e., being more sensitive about the aspects that the group mirror reflects) in the baseline condition, possibly leading to an increase in performance even when no group mirror is displayed.

Results show, furthermore, that the roles of the realist and the critic are used less often than the dreamer in the baseline. Group members already knew each other and their task was to discuss the ideas of their colleagues. Especially the role of the critic (and also the role of the realist) probably was less popular, as participants did not want to offend their colleagues by being critical about their ideas. With the group mirror, people stated more contributions matching the realist and the critic role. At the same time, disruption and stress did not increase, indicating that the more critic views were not accompanied with people feeling more stressed.

6.2 An Interactive Group Mirror for the Disney Method

In the previous section, I described a group mirror supporting the Disney Method using an unobtrusive and subtle visualization. In this section, a system designed to support the Disney Method is described that enables group members to interact with it directly. Two versions were compared in a laboratory study: a baseline in which the system is used without additional support and a condition in which the system additionally provides feedback about the balanced use of the different roles.

6.2.1 Background and Motivation

Equally to the group mirror described in the previous section (see Section 6.1), the system discussed in the following uses the concept of mirroring more content-related information (i.e., the distribution of the Disney Method's roles) instead of information identifying performance of individual group members. With this, we attempt to reduce pressure and stress induced by group mirrors. In particular, we are interested if such a visualization can still influence collaborative creativity in a positive way. Results of the field study described in the previous section indicate that more ideas are stated with such a group mirror, particularly in the roles that are normally less popular (i.e., the critic and the realist). To validate or revise the results from the field study, we present a larger laboratory study in this section.

Furthermore, we present a novel prototype that includes the possibility to enter ideas directly. By that, the group enters the information that is needed for showing the balance of the different roles by themselves, thus, a person assessing the contributions by listening to the discussion is not necessary. We are interested, if this approach is suitable for creating group mirror visualizations. We realized the prototype by developing an interactive system that runs on tablet computers.

6.2.2 Concept and Design

After having discussed some first ideas, we improved the most promising idea that was then used during the study. The design process is briefly described in the following.

Initial Design Considerations

The basic idea behind our prototype was to combine direct input of ideas by the group members with a visualization of the balance of the roles. One of our first ideas was to build on the system described in the previous section and to extend it with a functionality to add ideas that are displayed as colored text fields on the screen. The visualization of the amount of ideas matching the three roles is displayed in the background (see Figure 6.4,

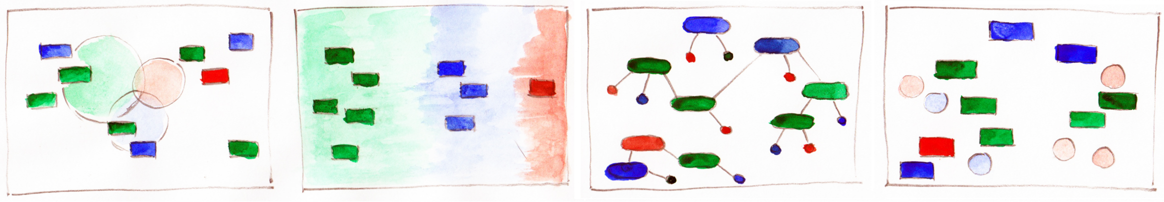


Figure 6.4: First sketches for the interactive system supporting the Disney Method. Left to right: 1. A combination of text fields and colored circles displays the amount of ideas of the roles. 2. Text fields are clustered dependent on their role, the background is colored accordingly 3. Each idea is connected to at least one of the other roles, unfilled text fields are indicated with smaller circles 4. Unfilled circles indicate underrepresented roles.

first sketch). Another idea was to predetermine some kind of clustering. The background is colored according to the amount of ideas that have been stated to that role. All ideas are collected in these areas (see Figure 6.4, second sketch).

However, both approaches have some disadvantages. Using the circles in the background can be problematic when too many ideas are displayed on the screen that occlude this information. The second approach restricts the ability to arrange and cluster ideas freely. It might, for example, make sense to cluster a dreamer contribution with a realistic and a critic viewpoint. This concept is reflected in the sketch in Figure 6.4, third sketch. Here, each idea of a specific role is connected to at least one idea of both other roles. Showing empty text fields ensures that group members get aware of underrepresented roles. However, this concept also restricts the arrangement, as it might not always be necessary to connect ideas of all three roles. We refined this idea and decided to leave out the connecting lines to make arrangements of ideas easier (see Figure 6.4, forth sketch). Now, group members can arrange ideas by locating them near each other. Underrepresented roles are displayed as empty circles. The final design is described in more detail in the following.

Final Design

Similar to the system described in Section 6.1, the system depicts the different roles of the Disney Method using different colors (dreamer: green, realist: blue, critic: red). Each group member is equipped with a tablet on which the system runs. The screen shows a menu bar and a canvas (see Figure 6.5). Group members can create and delete ideas by dragging a color from the menu bar. An empty circle appears. In the baseline, only the circle of this color shows up. In the condition with feedback mechanism, additional circles with the color of the remaining two roles turn up. With a double click on a circle, group members can enter an idea. For that, a keyboard appears on the screen. Figure 6.6 shows this procedure in both conditions.

On the screen, filled, unfilled and blocked ideas are visible. They can be moved and arranged freely on the canvas. Everyone has the same view on the interface. All actions are synchronized between the tablets. To enable parallel working, ideas that are currently altered

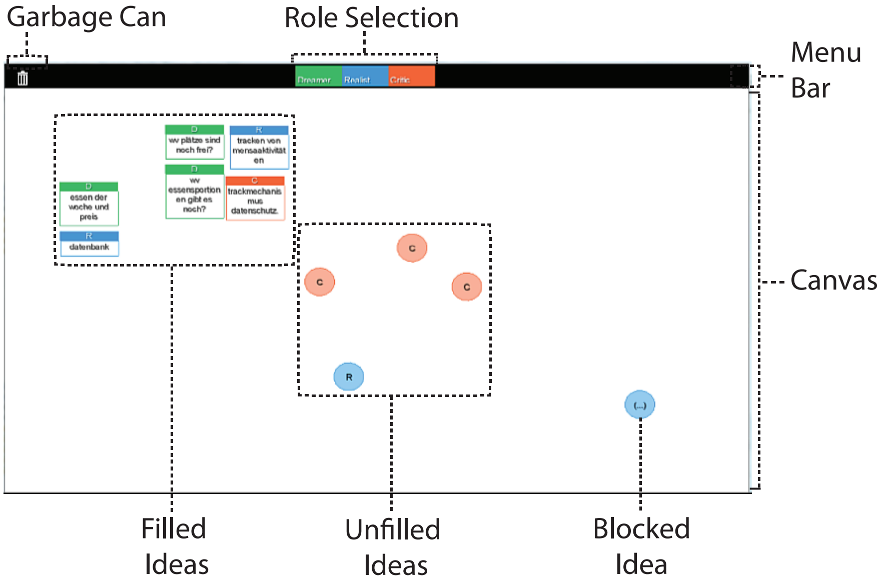


Figure 6.5: System design. The system is composed of a menu bar with which ideas can be created and deleted, and a canvas with filled, unfilled and blocked ideas.

by one person are displayed as “blocked” for all other group members (these circles display the characters “(...)”). Inside of the unfilled circles and on top of the ideas that are already filled, a character indicates the role (“D” for dreamer, “R” for realist, and “C” for critic). We used this information additionally to the colors to make the system accessible for color blind people. Filled ideas are displayed as rectangles containing the text while unfilled ideas are represented with circles. This should help recognizing which roles are underrepresented, as people perceptually group similar shapes together (Max Wertheimer, 1944).

The system was implemented in two versions. In the baseline version, people can create, alter and delete ideas as described before. In the feedback version, additional feedback is shown about which roles are underrepresented by displaying empty circles of the corresponding colors. When x, y and z are the amount of existing ideas of the three roles, the number of empty ideas of role x calculates as follows:

$$EmptyIdeas(x) = max(x, y, z) - x$$

The system is implemented as an Android app. To enable all group members to have a view, which is consistent with the views of the other group members, a client-server architecture is used. The clients run on the tablets, while the server holds all data. Communication is enabled over wireless network through a TCP socket.

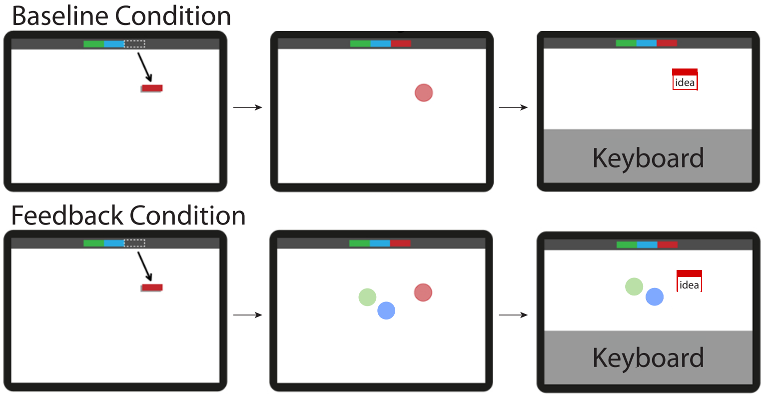


Figure 6.6: Functionalities of the system. To create an idea of a specific role, the corresponding color is dragged from the menu bar to the canvas. In the baseline, an empty circle of this color appears. In the feedback condition, empty circles of the two other colors appear additionally. The idea can be filled in by double-clicking on it.

6.2.3 Evaluation

We conducted a laboratory study to compare two versions of the system. In the baseline condition, the system was used without additional feedback about the balance of the roles while the feedback condition showed additional feedback with the goal to promote underrepresented roles. The main questions were, which effect this feedback has on the balance of the different roles of the Disney Method, how it affects collaborative processes and how the group accepts the system.

Method

We used a repeated measures, one-factorial design for our study. We compared two conditions, a baseline without feedback mechanism and a condition with additional feedback. Two topics were chosen. We counterbalanced condition $\times$ topic, thus, each unique combination was assigned to two groups.

Setup and Procedure

The study was conducted in a quiet room. In each session, three participants sat around a rectangular table. Each group member was equipped with a tablet (see Figure 6.7). We used three Samsung Galaxy Tab 3 tablets.

Before the study began, the experimenter gave a brief introduction to the study. Then, the Disney Method was introduced. Afterwards, the interface was explained. Participants could use the system until they understood how to use it properly. Two topics were used for the study that were given to the groups before each session: (A) *Ideas for an app for the students' cafeteria* and (B) *Ideas for an app for a dating platform*. Each group performed two rounds that lasted 15 minutes each. During the discussions, groups were asked to record their ideas

Figure 6.7: Study setup. Three group members using the interactive system (picture re-staged).

using the app. Each group member was allowed to switch roles dynamically . After the study, pen-and-paper questionnaires with 5-point Likert scales (1 = strongly disagree, 5 = strongly agree) and questions about demographics were handed out. Furthermore, short, informal interviews were held with the group.

The study was audio and video recorded. The system logged all activities of all group members with the system (i.e., creation, deletion and alteration of ideas with timestamp, participant ID and content).

Participants

In the study, 24 voluntary participants (8 female; average age 23, range: 18 to 29 years) took part in groups of three, 20 were students, 3 research assistants and one an employee in another professional field. In all 8 groups, participants knew each other before the study. All participants received a 10€ voucher from a well-known online web-store.

Results

We used a paired t-test for evaluating quantitative data. Standard deviations were compared to evaluate the balanced use of the roles and the balance of participation. The software Excel was used for the analysis. Qualitative data was gathered from the questionnaires and interviews.

Performance We analyzed the overall number of contributions per group by comparing them using a paired t-test. In the baseline, groups produced slightly less ideas (M = 22.13, SD = 6.38) compared to the condition with feedback mechanism (M = 25.13, SD = 11.18). The t-test did not reveal a significant difference between the two conditions.

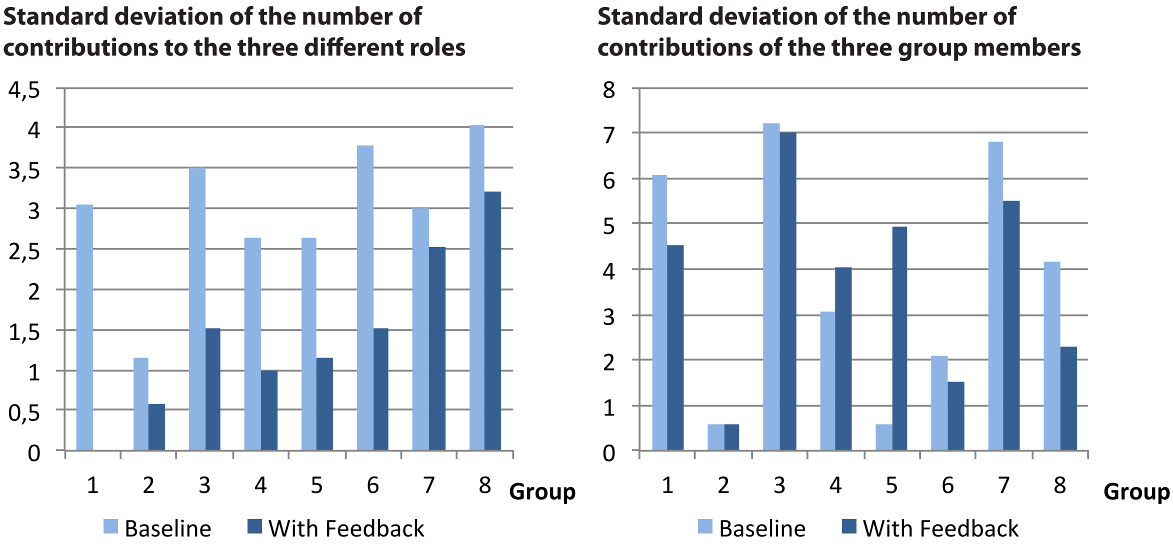


Figure 6.8: Results on balancing of the roles and participation. Left: Standard deviations of number of ideas recorded for the three different roles for all eight groups. Right: Standard deviations for the amount of ideas entered from the three participants for all eight groups.

Balanced Use of the Roles The previous study indicated that the "unpopular" roles, such as critic and realist, are used more often with the feedback of a group mirror. We were interested, if the same effect can be observed in this study with the interactive feedback system. Dreamer ideas were recorded similarly often in the baseline condition (M = 9.5, SD = 3.7) and the feedback condition (M = 9.5, SD = 3.25). The roles of the realist and the critic were used more often with the feedback mechanism than without (Realist in the baseline (M = 7.63, SD = 2.97) and with feedback (M = 8.5, SD = 3.66); Critic in the baseline (M = 5.13, SD = 2.36) and with feedback (M = 7.13, SD = 4.26)). However, a paired t-test did not show significant differences. Still, these results indicate that ideas of the realist and critic roles were not stated instead of dreamer ideas, but additionally to these. To take a closer look at how balanced the roles were used we calculated the standard deviation between the three roles. Figure 6.8 (left) shows the standard deviation in the baseline and with feedback for all groups. This indicates that with the feedback mechanism, the roles are used in a more balanced way.

Balanced Participation We were, furthermore, interested, how persons with different activity levels were affected by the feedback. Therefore, we categorized participants into *below average* and *above average*, similarly to the study with GROUPGARDEN (see Section 5.1). Again, this categorization was performed after the study so that the group members did not know about their categorization. We took the baseline as a basis and divided the mean number of ideas per group by three to get the mean number of ideas per participant (M = 7.71). Group members with more ideas than that were categorized as *above average*, participants with less as *below average*. This resulted in 10 *above average* and 14 *below average* participants.

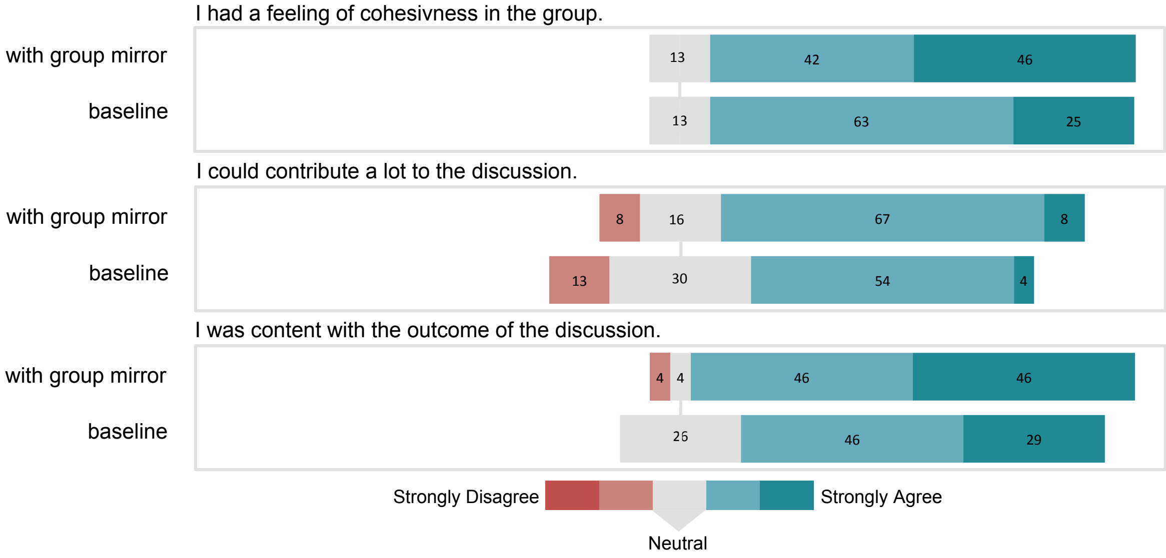


Figure 6.9: Results of the questionnaires. Answers to the questions about cohesiveness, feeling of the amount of contributions and about the satisfaction with the discussion. Numbers indicate the percentage of participants who answered with that score on the 5-Point Likert scale. Numbers are rounded and thus might not add up to 100% exactly.

Participants categorized as *above average* slightly increased the amount of ideas with the feedback (M = 12.8, SD = 3.65) compared to the baseline (M = 12.1, SD = 2.6). This is an increase of 5.8%. *Below average* participants increased the number of ideas more. With feedback mechanism, 32,8% more ideas were generated (M = 6.07, SD = 4.13) than in the baseline (M = 4.57, SD = 1.83). We calculated a t-test for both of these groups, however, differences are not statistically significant.

Additionally, we calculated the standard deviation of the amount of ideas of the group members. Figure 6.8 (right) shows the results for all eight groups. For five groups, the standard deviation was smaller with the feedback system, for one it was the same, for two groups the standard deviation was bigger. This indicates that for five of eight groups, participation was more balanced with the feedback mechanism.

Perceived Effects on Group Processes We handed out questionnaires after both conditions to compare the perceived effects on group process. For instance, we asked how cohesive the group was perceived. Figure 6.9 (top) shows the results. With group mirror, 11 participants (46%) strongly agree that the group work felt cohesive, 10 participants (42%) agree. In the baseline, 6 participants (25%) strongly agree, 15 (63%) agree. In both conditions, 3 participants (13%) are neutral regarding this aspect. This shows a slight tendency to a better feeling of cohesiveness when the feedback mechanism is shown. This also applies to the feeling of being able to contribute a lot to the discussion (see Figure 6.9 (middle)). In the condition with feedback, 2 participants (8%) strongly agreed, 16 participants (67%) agreed, 4 participants (16%) were neutral and 2 participants (8%) disagreed. In the baseline, 1 participant (4%) strongly agreed, 13 participants (54%) agreed, 7 participants (30%) were neutral and 3 participants (13%) disagreed. Finally, we asked if people were content with the outcome of the discussion (see Figure 6.9 (right)). With group mirror, 11 participants (46%) strongly agreed, another 11 participants (46%) agreed, 1 participant (4%) was neutral and 1. participants (4%) did not agree. In the baseline, 7 participants (29%) strongly agreed, 11 participants (46%) agreed and 6 participants (25%) were neutral. Again, a slight tendency to more people feeling more content with feedback mechanism is visible.

Acceptance of the System In the questionnaires and the interviews, we asked participants about their thoughts about the system in general. 23 of 24 participants valued the synchronicity of the interface. While for some it was confusing in the beginning, all participants found it helpful after familiarization. Participants of group 1 appreciated that they could enter ideas synchronously, since it reduces phases where one cannot contribute as others are talking (i.e., reducing production blocking). Moreover, we asked participants in the questionnaires to comment on what they liked or did not like. Seven participants would have liked to have the option to connect ideas, as they are used to this in other, similar affinity diagramming tools. Five participants mentioned that double clicking on empty contributions to enter ideas was an unfamiliar gesture for touch devices. However, nearly half of the participants (11 participants) perceived interaction with the system as *"easy"* or *"intuitive"*. When asked about the feedback mechanism, 15 participants reported to perceive it as *"annoying"* or *"disrupting"*, as new empty circles appeared in the middle of the screen. However, others (4 participants) explicitly stated that it was *"motivating"* for them to see the suggested roles.

Summary and Discussion

Building on the prototype presented in the previous section, we designed an interactive system supporting the Disney Method. We compared two versions, one that allows groups to enter, change and delete ideas using the three roles of the Disney Method (dreamer, realist, critic) and one that additionally gives hints on currently underrepresented roles.

The results indicate that more ideas were generated with the feedback mechanism than without. Especially *below average* participants increased the number of ideas with the feedback (however, without statistically significant differences). Several reasons are imaginable. First, the effect of free riding might be reduced through the guidance of the feedback. Free riding names the effect that people rest on the effort of the other group members, thinking that their contributions are dispensable (for a more detailed discussion on free riding see Section 3.1.2). The feedback provides information about missing roles. This might lead to the effect that group members notice that their ideas are not dispensable but helpful and desired. Second, this form of guidance might assist group members who do not have much experience in creative idea generation. Just discussing about a certain topic might be too vague and could be the reason for creating ideas merely in a certain direction. The guiding feedback might help to think about a topic from different angles and to include more diverse viewpoints. Additionally, results indicate that "unpopular" roles such as the critic and the realist are used more often with the feedback mechanism.

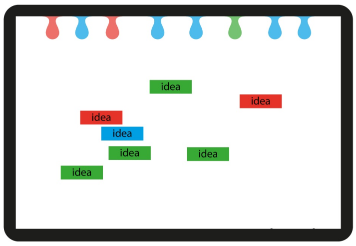


Figure 6.10: Idea for future work. The feedback about underrepresented roles could be displayed in the menu bar. By dragging the drops to the canvas, an idea of this role is created.

However, the feedback about the underrepresented roles was at the same time perceived as annoying and disrupting. More unobtrusive forms of feedback could be one solution to this problem. For instance, the menu bar with which ideas are created could be used to provide the information about underrepresented roles without limiting the screen space too much. Figure 6.10 shows, how this could look like. The number of available drops in the menu bar implicitly suggests, which roles have been utilized less than others.

6.3 Chapter Summary

In this chapter, two prototypes were discussed that pursued the goal to support the Disney Method, a role-based collaborative creativity technique. The first prototype was a subtle visualization displayed on a tabletop display, showing information about the role distribution. The second prototype is a more interactive system enabling group members to create ideas on tablet computers. We conducted a field study and one laboratory study. The main results are listed in the following:

- Showing groups a visualization of the distribution of the different roles of a role-based creativity technique can increase motivation and lead to more productive idea generation.
- Visualizing role distribution can lead to more balanced role distribution. Especially "unpopular" roles such as the critic or the realist are used more often with support of a visualization, since group members are reminded of the different roles and the frequency of use of these roles.
- There might be similar positive effects of group mirrors focusing on content-related feedback (i.e. the feedback of the distribution of the different roles) regarding performance and balance of participation compared to group mirrors showing more person-related information (i.e., the performance of the individual group members).
- The reported results of the second study express merely tendencies. None of the quantitative results revealed statistically significant differences. One could assume that this

is due to the characteristic of the feedback system that shows content-related feedback. However, the first project on the Disney Method (see Chapter 6.1) also showed content-related information and still, an increase of performance with group mirror could be observed. It might be possible that in the second study, the baseline and the feedback condition were too similar to lead to significant differences. The feedback system of the first study was compared to a baseline without any technological support, in contrast to the second study, in which two versions of a system were compared. Both showed the aggregated performance of the whole group. Only imbalances in the role distribution were highlighted in the feedback condition. Thus, this feature alone may have a weaker effect compared to the feedback about the performance of the group. However, this is just an assumption and would need further investigations (i.e., by comparing content-relating feedback without an indication of the group performance to a baseline without feedback). For a general discussion on the issue of significance testing, see Section 9.2.5.

IV

SUPPORTING ARGUMENTATION WITH GROUP MIRRORS

7
Supporting Collaborative Argumentation

After having elucidated the impact of several group mirrors on creative tasks, the following chapters will focus on two other use cases, argumentation and debates. As described in the design space chapter (see Chapter 4), group mirrors are applicable in a variety of different tasks. These tasks can be regarded as a continuum from more open-ended to more structured tasks. We chose tasks on this continuum that can be seen as representatives for more open-ended tasks (the collaborative creativity techniques discussed in the previous chapters) and more structured tasks (argumentation and debates as discussed in the current and the following chapter).

Moreover, argumentation itself in its different manifestations can be more or less structured. Collaborative argumentation as approached in this thesis can be seen as more open-ended compared to debates practiced in debate clubs, which - in contrast to our argumentation tasks - specify a particular order of the speakers and elect a winning team after the debate. In this chapter, a system supporting collaborative argumentation along with two studies will be described. Systems supporting more structured debates will be addressed in Chapter 8.

7.1 Study I: A Comparison of Anonymous and Identifiable Feedback

This section presents a prototype that enables subtle peer feedback during argumentative debates. In a study, we compared three conditions, a baseline without any support through a tool, an *anonymous* version of the system and an *identifiable* version. The motivation, design and the results of the study will be presented hereafter.

This chapter is based on a master thesis (Reithmeier, 2013) and a bachelor thesis (Attwenger, 2015). The detailed personal contribution statement can be found in the disclaimer.

7.1.1 Background and Motivation

Our main goal with the prototype presented in this chapter is to support argumentation by helping people in learning how to build proper arguments. Argumentation and discussions are ways to exchange knowledge, to learn from each other, to solve problems and to come to decisions. However, many people are poor arguers (Tannen, 1998), so learning how to argue is the topic of various research studies. Discussions hold inherent challenges that need to be addressed to make them beneficial (Doyle and Straus, 1976). Research in the field of Learning Sciences and Psychology has shown that discussions can benefit from involving all participants (Burns, 1995), by linking arguments to each other (Leitão, 2000) and by using well-formed arguments (see e.g., Rieke et al., 2009), for instance, by using the elements *claim, ground* and *qualification* (Toulmin, 1958; Weinberger et al., 2007). Building on that, tools have been developed that support the learning of these concepts (for an overview see Scheuer et al., 2010). Some approaches investigated the effect of immediate feedback on argumentation skills (e.g., Goodman et al., 2005; Israel and Aiken, 2007; Shute, 2008), however in contrast to the work presented in this chapter solely in the context of remote computer-supported argumentation. Chapter 3.2 shows a more detailed explanation of the benefits and challenges of discussions and insights of related research.

Our approach to help people to learn how to argue makes use of providing and receiving peer feedback about the argument structure. Using the assessment of peers has several advantages. On the one hand, rating the quality of an argument depending on its structure in real-time is yet difficult to complete for computer systems. On the other hand, peers can not only learn from receiving feedback but also from providing feedback. There is reasonable evidence that peer feedback can improve learning outcomes since people reflect on their own and their peers' performance (Cho and Schunn, 2007; Loll and Pinkwart, 2009).

We developed a tool consisting of physical light cylinders that displays the feedback of the peers to the group members. Verbal feedback from a person has the drawback that it either interrupts the discussion or can only be provided after the discussion. In contrast, real-time feedback by peers, mediated through a group mirror, does not require verbal interruptions. Using a technical medium for providing feedback also makes it possible to conceal the identity of the originator of the feedback.

In the user study outlined in the following, we compared two versions of the light cylinders, one using identifiable and one anonymous feedback, to a baseline. Anonymity in debating has been investigated in the domains of group decision making or social psychology. Anonymity can be defined as *"the inability of others to identify an individual or for others to identify one's self"* (Christopherson, 2007). Both positive and negative effects of anonymity in debating have been described (see e.g., Christopherson, 2007; Dubrovsky et al., 1991; Nunamaker et al., 1996; Sia et al., 2002). For example, a study conducted by Flinn and Maurer (1996) suggests that more critical feedback is provided and performance increases with anonymous feedback. Ainsworth et al. (2011) investigated anonymity in classroom debating and voting. The authors could show that students voting anonymously were less likely to vote in alignment to the group norms. This might be attributable to the requirement

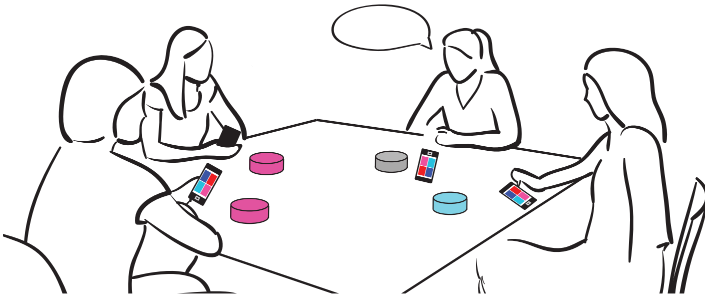


Figure 7.1: Functionality of the light cylinders. The light cylinders are controlled via smartphone and are used to provide feedback to the current speaker. In this picture, feedback is identifiable as the speaker can see who gives which feedback.

of unveiling their view to the other students in a non-anonymous situation rather than being exposed to the views of the other students. Results of their studies, furthermore, indicate that anonymity in debating (in their case by using CMC) had some negative effects, for instance, an increase in off-task behavior.

However, these studies either investigated anonymous debates, meaning that the members of a debate use a system to enter their contributions, or anonymous voting, which was conducted after the debate. In our case, not the arguments themselves but the quality ratings of arguments are provided anonymously in real-time, in a co-located scenario. A similar system has been presented by Bergstrom and Karahalios (2007b). Their tool allowed providing anonymous votes on discussions (see also chapter 2.2.2). It promoted especially active and unsatisfied voters to participate more, leading to more diverse opinions being included.

The differences between *anonymous* and *identifiable* feedback in co-located debates have, to our best knowledge, not been investigated yet. In the study presented in the following, we focus on effects of the identifiability or anonymity of the feedback. Identifiable feedback reveals its originator who can add social cues to the mediated feedback (facial expression, gestures), whereas anonymous feedback cannot be traced back and therefore misses the quality of social cues. To investigate the effects on performance and social acceptance, we used a group mirror system in form of light objects that enable participants to provide real-time feedback to each other in a co-located discussion (see Figure 7.1). Group members have the possibility to indicate agreement and disagreement and provide feedback about the structure of arguments.

7.1.2 Concept and Design

To compare the influence of identifiable and anonymous feedback we built a group mirror that can be used for both conditions. We designed light objects inspired by LANTERN (Alavi and Dillenbourg, 2012) and STATUBE (Hausen et al., 2012). We adopted the idea of different levels, so that each level can also be used as stand-alone system. Light colors can be controlled via smartphones (see Figure 7.1).

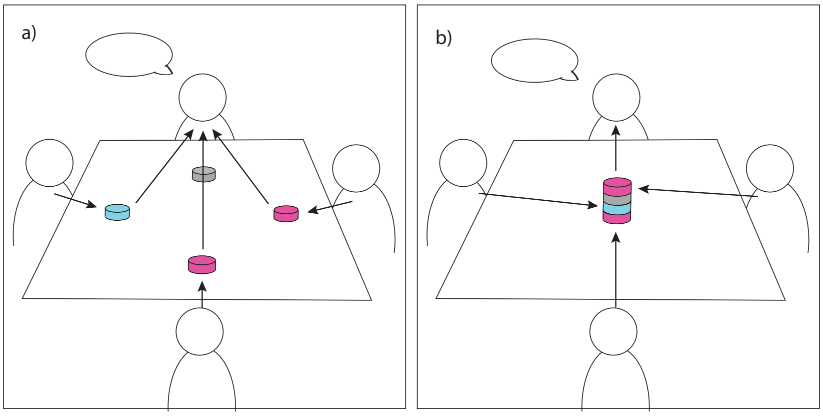


Figure 7.2: Study setup. a) Light cylinders are placed in front of each group member to provide identifiable feedback to the current speaker. b) Light cylinders are stacked on top of each other in the middle of the table and feedback is sent to a randomly selected cylinder to provide anonymous feedback.

In our scenario, each participant is equipped with a smartphone and a light cylinder and can provide feedback to the current speaker. The cylinders are placed on a table so that every participant can see them. When each participant has a personal light cylinder that stands in front of him, the feedback is identifiable, as the current speaker is able to recognize who gave which feedback (see Figure 7.2a). When the light cylinders are stacked on top of each other in the middle of the table, feedback is provided anonymously. The actual object and its location on the stack is chosen randomly, so that the origin of the feedback remains hidden. (see Figure 7.2b).

We are interested in feedback on the argument structure. However, in a pilot study, participants remarked that they sometimes used the feedback mechanism to indicate agreement or disagreement with an argument instead. It seems to be difficult for people to separate these two types of feedback, when they are not clearly made aware of this difference. As a result, we decided to provide both possibilities, stating agreement or disagreement and rating the structure of the argument. An argument is structured well when it consists of the elements *claim, ground* and *qualification* (Toulmin, 1958; Weinberger et al., 2007). We call arguments that fulfill these requirements *well-structured*. Arguments that lack either ground or qualification are called *unsupported claims*. One example of a well-formed argument taken from the records of the study (with the topic "Should opening hours of shops be extended?") is: *"Yes, I think that this will lead to rising prices [claim]. When you extend opening hours, you have to employ people longer. (...) Costs more power, more working hours [ground]. The question is, if the earnings of the shops are big enough [qualification]."*

The smartphone app consists of four buttons to provide the feedback (see Figure 7.3). Blue indicates agreement to an argument, red disagreement, cyan that the argument is well-formed and magenta is used for unsupported claims. Combinations of these four categories (such as disagreement to a well-formed argument) were excluded for reasons of clarity and memora-

Figure 7.3: Light cylinders. The cylinders were built of semi-transparent acrylic and equipped with an Arduino board and LEDs. Colors of individual cylinders are changed via a smartphone app.

bility. A card was placed in front of each participant to remind him or her about the meaning of the colors. We intentionally did not include this information in the smartphone app, as participants would then have to look down to the smartphone, which would interfere with the anonymity of the feedback in the anonymous version.

To maintain anonymity, the buttons of the smartphone app need to be pressed in an unobservable way, e.g., by holding the smartphone under the table. We equipped smartphones with a soft rubber cover with four cuttings positioned over the buttons so that the feedback could be provided blindly. All of the participants stated after the study that they could not detect who gave which feedback.

Hardware A semi-transparent acrylic tube was cut into pieces to create the cases of the cylinders (see Figure 7.3). One cylinder is 4cm high and has a diameter of 9cm. The lid and the bottom of each cylinder are provided with a coupling mechanism, so that they can be stacked on top of each other to build one big cylinder. Each cylinder is equipped with a circular array of LEDs to illuminate the cylinder. An Arduino board[1] with Bluetooth module is used to control the light. A Bluetooth connection is used because of the simple establishment of connections and because only a small range was required. A wireless power supply is assured via a 9V battery.

Software The Arduino board was programmed to light up the LEDs for ten seconds after a Bluetooth signal was received. An Android application with the user interface described above communicates with the system via Bluetooth and Wi-Fi. When the system is used as identifiable feedback, each group member is provided with a smartphone and a cylinder. The smartphone connects to the personal cylinder and sends signals via Bluetooth to change the color of the cylinder. To achieve anonymous feedback, a computer is interposed between smartphones and cylinders. The computer receives signals from the smartphones via Wi-Fi and sends them as Bluetooth signals to a randomly chosen cylinder. The cylinder lights up for ten seconds. Only within that time, the cylinder is associated to that specific smartphone. For a next signal, a new cylinder is chosen randomly.

[1] www.arduino.cc, last accessed 08.01.2016

7.1.3 Evaluation

Our main goal was to investigate the effects on group processes achieved by decoupling feedback content from information about the feedback provider. We therefore conducted a study in which we compared an anonymous group mirror hiding the identity of the feedback originator to an identifiable version showing the feedback originator and to a baseline.

Method

A between-group design was chosen for the study. We used a one-factorial design with three conditions. We varied the *level of identifiability*. The three conditions are: (1) groups without mediated feedback (baseline) (2) groups with anonymous feedback and (3) groups with identifiable feedback. The between group design was chosen to limit the time commitment for participants. Each group took part in one of the three conditions. In each condition, groups discussed two topics. Half of the groups discussed topic A at first, the other half topic B.

Setup and Procedure

Participants took part in the study in groups of four and were asked to sit around a rectangular table (see Figure 7.4). As the study was conducted between groups, each group only participated in one condition. In the beginning, a short introduction to the procedure of the study was given. First, the structure of well-formed arguments was explained. An example argument relying on the model of Weinberger et al. (2007) was given. After that, the groups with feedback systems received instructions on how to use the group mirror. Before each session, one of the two topics was introduced. The topics were: A) *Is speed limit on freeways reasonable?* and B) *Should opening hours of shops be extended?* These topics were chosen because arguments exist for pro and contra sides. Moreover, these topics are common in local political debates and the majority of people know about these issues. Group members were asked to build well-formed arguments, composed of the parts *claim, ground* and *qualification* (Toulmin, 1958; Weinberger et al., 2007).

Figure 7.4: Study setup. Left: Identifiable condition. Right: Anonymous condition (pictures re-staged).

A session consisted of two discussions taking both approximately ten minutes. The three conditions were:

Baseline In the baseline, groups did not provide and get any mediated feedback. Excluding feedback in the baseline was necessary to allow comparison of both conditions to the baseline. If feedback would have been provided by a facilitator in the baseline, a comparison to the identifiable condition would have been possible. However, a comparison between the anonymous condition and the baseline would have changed two variables, the identifiability of the feedback (anonymous vs. identifiable) and the feedback origin (person vs. system).

Identifiable Feedback As in the anonymous condition, participants could provide feedback about agreement, disagreement and argument structure to the current speaker. Each group member was equipped with a smartphone and could transmit feedback to the light cylinder in front of him or her (see Figure 7.2a). To avoid additional constraints, verbal feedback was not explicitly forbidden.

Anonymous Feedback Groups could provide feedback about agreement and argument structure. All cylinders were stacked upon each other in the middle of the table. Participants were asked to place their smartphones under the table so that the input was not observable by the others (see Figure 7.2b). Again, verbal feedback was not explicitly forbidden.

Each discussion was followed by a questionnaire, so that each participant filled out the same questionnaire twice. Finally, we conducted a semi-structured interview about the discussion and the feedback system. All participants were thoroughly debriefed.

Data Collection

We collected quantitative as well as qualitative data. All sessions were audio and video recorded, discussions as well as interviews were fully transcribed and input data on the smartphones was logged. SPSS Statistics was used for analyzing the quantitative data. As described above, four participants formed one group. We therefore needed an analysis that takes into account that participants are nested in groups. We therefore performed a nested ANOVA test.

Quantity of Feedback To collect data on how often the different types of feedback (agreement/disagreement and rating about the formal structure of an argument) were used, the smartphone app logged all input during the discussion.

Quality of Argumentation In order to objectively assess the quality of arguments, we recruited two independent experts. The experts had both been members of a debate club for four years and had considerable experience in rating argumentation. The transcriptions of the discussions were handed out to the experts without indicating the condition in which the discussion took place. The experts defined three types of arguments. The first type is arguments with claim that are emphasized by a very detailed ground, a general view on the topic and an example. The second type is arguments consisting of claim, short ground and short qualification. The third type contains a claim and a short ground but lacks a qualification. Claims without any explanation were not counted as an argument. As the first type of arguments was rare, we summarized the first two categories. We call these arguments *well-formed*, whereas arguments without qualification are called *unsupported claims*.

Quantity of Arguments We used experts' evaluation results to count well-formed arguments and unsupported claims.

Number of Interruptions We used video recordings to measure how often participants interrupted each other. Any occurrence of two or more people talking simultaneously was counted as interruption. In this way we wanted to investigate, how much the type of feedback influences groups to observe general discussion rules such as not to interrupt each other.

Perception of Feedback Questionnaires with 5-point-Likert scales (1 = strongly disagree, 5 = strongly agree) were handed out after each discussion to get information about the participants' subjective perception of the feedback systems. At the end of each session we conducted a semi-structured interview with the group. We conducted a focused coding approach to analyze these interviews. Codings were then categorized and discussed collaboratively.

Participants

We recruited 48 voluntary participants who took part in the experiment in groups of four. We deliberately included a broad variety of different ages (average age 37, range: 19 to 76) and professions. Most of the participants were employed (28) or students (13), others were trainees or retired. We counterbalanced female and male participants (50%). The members of each group already knew each other before the study. All participants were reimbursed with a 10€ voucher.

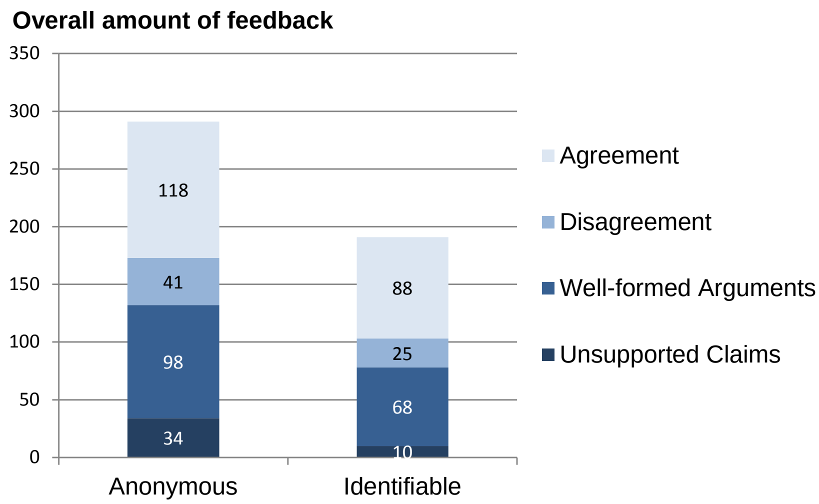


Figure 7.5: Results on the amount of feedback. A tendency is visible that the overall amount of feedback provided in the anonymous condition was higher than the amount provided in the identifiable condition.

	Condition	Mean	SD
Overall Feedback	Anonymous	18.06	6.913
	Identifiable	11.81	4.246
Agreement	Anonymous	7.35	3.768
	Identifiable	5.50	2.582
Disagreement	Anonymous	2.56	1.315
	Identifiable	1.69	1.815
Well-formed Arguments	Anonymous	6.13	2.872
	Identifiable	4.25	2.852
Unsupported Claims	Anonymous	2.13	2.473
	Identifiable	0.63	1.088

Table 7.1: Results on the amount of feedback. Mean values and standard deviation (SD) of the number of different feedback information in the two conditions.

Quantitative Results

We evaluated the quantity of feedback, the quantity and quality of arguments (i.e., the compliance to the structure of arguments) and the number of interruptions.

Quantity of Feedback We conducted a nested ANOVA on the two feedback conditions to compare the quantity of different feedback categories (agreement and disagreement to an argument and formally complete or incomplete argument). The logging data indicates that group members in the anonymous condition used the feedback system more often (M = 18.06, SD = 6.913) in comparison to groups with identifiable feedback (M = 11.81, SD = 4.246), without statistical significance ($F(1, 6) = 5.771$, $p = .053$) (see Figure 7.5). Table 7.1 depicts mean (M) and standard deviation (SD) for the different feedback types.

Figure 7.5 shows the overall amount of the different types of feedback that were provided. Feedback about well-formed arguments and agreement was provided more often in the identifiable condition: In the anonymous condition, 40.5% (absolute number 118) of the feedback indicated agreement and 33.7% (98) indicated that an argument was well-formed in comparison to 46.1% (88) indicating agreement and 35.6% (68) indicating a well-formed argument in the identifiable condition. Accordingly, less feedback about unsupported claims and disagreement was provided in the identifiable condition. In the anonymous condition, 14.1% (41) of the feedback indicated disagreement and 11.7% (34) indicated unsupported claims while in the identifiable condition, 13.1% (25) of the feedback indicated disagreement and 5.2% (10) indicated unsupported claims.

Taken together, feedback in the categories agreement and well-formed arguments was provided more often in the identifiable condition (M = 13.5, SD = 5.704) than in the anonymous condition (M = 9.75, SD = 4.235). Feedback in the categories disagreement and unsupported claims was given more frequently in the anonymous condition (M = 4.69, SD = 3.361) than in the identifiable condition (M = 2.31, SD = 1.887). A nested ANOVA did not reveal a significant difference.

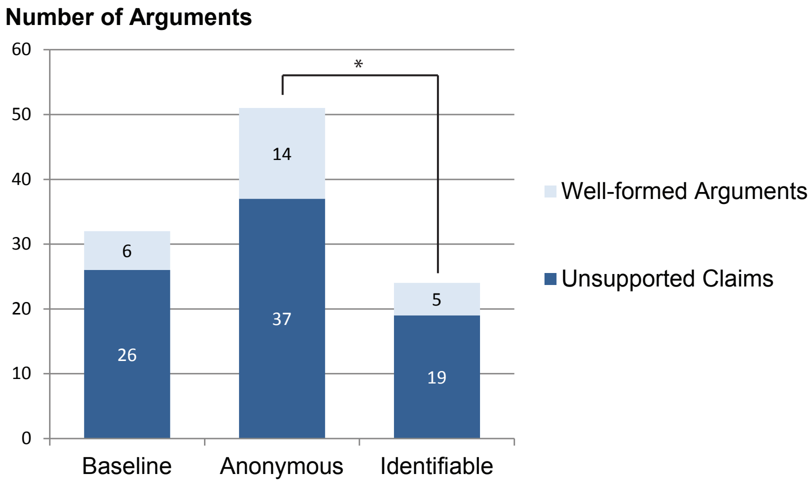


Figure 7.6: Results on the quality of arguments. Two independent experts rated the arguments in the categories *well-formed arguments* and *unsupported claims*. In the anonymous condition, more arguments and more well-formed arguments were produced.

Quality of Argumentation The main intention of the prototype was to increase the quality of the argumentation. Results show that with the anonymous version, the quality of the arguments increased significantly compared to the identifiable condition. To evaluate the quality of the arguments, the two experts used a slightly adapted evaluation standard that normally is used in professional debate matches. They rated the majority of arguments as unsupported claims. The reason for this may be the varying discussion experience of the experts and the study participants. Nevertheless, trends in the distribution of arguments of high and low quality became apparent.

In the anonymous condition, 27.5% (absolute number: 14) were well-formed arguments in comparison to only 18.8% (6) in the baseline condition and 20.8% (5) in the identifiable condition. The experts rated 72.5% (37) of the arguments in the anonymous condition as unsupported claims in comparison to 81.2% (26) in the baseline and 79.2% (19) in the identifiable condition (see Figure 7.6). The results of the nested ANOVA show that the condition significantly affected the number of well-formed arguments ($F(2, 9) = 5.09$, $p < .05$). A post hoc Tukey test shows significant differences between the anonymous ($M = 0.88$, $SD = 0.719$ args.) and the identifiable condition ($M = 0.31$, $SD = 0.479$ args., $p < .05$) (see Table 7.2).

Quantity of Arguments Results, moreover, show an increase of the amount of arguments in the anonymous compared to the identifiable condition. We conducted a nested ANOVA on the three conditions to compare the quantity of the arguments. We were able to measure a significant effect of the different levels of anonymity on the quantity of arguments ($F(2, 9) = 5.23$, $p < .05$). In the anonymous condition, 51 arguments were built in total. In the baseline, 32 arguments were stated and in the condition with identifiable feedback 24 arguments were stated. A Tukey post-hoc test reveals a significant difference between the anonymous ($M = 3.19$, $SD = 1.834$ args.) and the identifiable condition ($M = 1.5$, $SD = 1.366$ args., $p < .05$).

	Condition	Mean	SD	
Number of Arguments	Baseline	2.00	1.826	
	Anonymous	3.19	1.834	*
	Identifiable	1.50	1.366	
Well-formed Arguments	Baseline	0.38	0.619	
	Anonymous	0.88	0.719	*
	Identifiable	0.31	0.479	
Unsupported Claims	Baseline	1.63	1.586	
	Anonymous	2.31	1.448	
	Identifiable	1.19	1.167	

Table 7.2: Results on the quality of arguments. Mean values and standard deviation (SD) of high and low quality arguments in the three conditions.

Number of Interruptions We conducted a nested ANOVA on the three conditions to compare the number of interruptions. We could not observe that the condition affected the number of interruptions significantly, but there is a tendency towards fewer interruptions when using the feedback system compared to the baseline (see Table 7.3).

Perception of Feedback In the questionnaires, we asked participants which kind of feedback (agreement/disagreement or rating of the formal structure) was more helpful. We will summarize the points *strongly agree* and *agree* under *agreement* and *strongly disagree* and *disagree* under *disagreement*.

Results show that in the identifiable condition, 13 participants (41%) did not perceive any difference. 13 participants (41%) valued the feedback about the agreement to an argument the most, 6 (18%) valued the rating about the formal structure of the argument the most. In the anonymous condition, 20 participants (63%) did not perceive a difference. 10 (31%) liked feedback about agreement best and 2 (6%) liked feedback about the formal structure the most.

Moreover, we asked participants how disrupting they perceived the feedback in form of the light cylinders. In the identifiable feedback condition, 13 participants (41%) agreed that the feedback system was disrupting, 6 (18%) were neutral regarding this aspect and 13 participants (41%) disagreed. In the anonymous mode in which the feedback system was positioned in the center of the table, participants seemed to estimate the feedback system slightly more disrupting. 14 participants (44%) agreed that the feedback system was disrupting, again 6 (18%) were neutral regarding this aspect and 12 participants (48%) disagreed.

Another form of disruption might derive from the use of the smartphone. We therefore asked the groups, how disrupting they perceived the interaction with the smartphone. In the identifiable feedback condition group members could easily look at their smartphone whereas in the anonymous condition, group members had to provide the feedback in an unobservable way. In the identifiable feedback condition, 9 participants (28%) agreed that interacting with

	Condition	Mean	SD
	Baseline	10.44	11.171
Interruptions	Anonymous	7.44	5.366
	Identifiable	6.06	5.039

Table 7.3: Results on the number of interruptions. Mean values and standard deviation (SD) of the number of interruptions in the three conditions

the smartphone was disrupting, 5 (16%) stated they were neutral and 18 participants (56%) disagreed. In the anonymous condition group members seemed to perceive the interaction with the smartphone more disrupting. 14 participants (44%) agreed that it was disrupting, 4 (12%) were neutral regarding this aspect and 14 participants (44%) disagreed.

Qualitative Results

After the session, we conducted semi-structured interviews with the groups, covering the attendees' experiences with following argumentation rules and the group mirror system they had used. Afterwards, we explained the other types of feedback to the group. Groups in the baseline were informed about both types of feedback. We then let them discuss advantages and disadvantages of the different feedback types. We will indicate groups from the baseline with the abbreviation B#, from the anonymous condition with A# and from the identifiable with I#. P# stands for the participant while # indicates the group's or participant's number.

General Concerns

In all conditions, group members expressed concerns about the use of our group mirror. Most positive remarks were stated in the identifiable condition. We will describe general concerns first and then present which advantages participants perceived when using the identifiable group mirror.

In the briefing we told participants how to build well-formed arguments. We did not predetermine any specific scenario. This probably led members to think about such a group mirror with respect to different scenarios in which they can imagine that such a tool is useful. They reflected about discussions in meetings, in job interviews or discussions with a learning background and seemed to find group mirrors especially useful for learning scenarios, which fits to our intentions for the system.

Disruption When participants talked about general discussions, they placed a bigger emphasis on the content of the arguments than on learning how to structure them. This might have influenced participants' judgment on how disrupting they perceived the system. *"I found it difficult to follow the discussion, not only to be concerned with the decision if you agree or disagree (...). Forming an opinion, then communicating the feedback and at the same time collecting my own answers and arguments, that was a bit too much and distracted me."* (I2P1). Others forgot to pay attention to the feedback altogether: *"At the first [discussion] I totally forgot to pay attention [to the feedback]. The second time I was concentrated so that I would perceive it, but it was... it was there."* (A3P1).

Participants that found the group mirror disrupting explained that with the additional cognitive load they needed when giving feedback. Other participants stated that they did not find the group mirror disrupting and experienced the small overhead of perceiving the feedback as justified: *"It doesn't disrupt too much. I mean [...] audio feedback would be more disrupting. In face-to-face discussions in general, the eyes are not very busy except for assessing the mimic of the others. It seems feasible to perceive the color pattern."* (I4P4).

Artificiality However, other concerns do not seem to be specifically problematic in discussions but can occur in learning situations, too. Group members were concerned about the naturalness of the situation and mentioned that the discussion felt artificial. One participant explained: *"I think it comes at the expense of naturalness, independent of the situation that is a bit unnatural anyway. I think because of the machines and you are concentrated and have to look at it all the time and that comes a bit at the expense of naturalness."* (A1P3).

Bullying As the system allows people to provide feedback to each other, some of the participants were afraid of bullying. They stated: *"I think, it is difficult that it doesn't turn into bullying. When you have a meeting and there is someone in it who always has weird ideas, he would always be voted down when he starts talking."* (A2P1). Especially the anonymity of the anonymous version might have evoked parallels to cyberbullying (Cassidy et al., 2013), as it combines a face-to-face situation with anonymity of the feedback. However, these concerns seemed to be based on assumptions rather than actual experiences. None of the participants recalled occurrences during the study, neither did we observe any.

Streamlining Participants also mentioned concerns of feeling intimidated and as a consequence suppressing their actual opinion. Especially if opinions differ from the common consensus, streamlining of opinions could occur: *"Some people could feel intimidated. They start to speak and then it says agreement or disagreement, your argumentation is bad."* (A2P2); *"I found that it was structured in a way so that extreme opinions would perhaps be blocked."* (A4P4); Again, these concerns were hypothetical, none of the participants actually mentioned feeling intimidated in the study situation.

Pressure Some participants thought about how the group mirror would be applied in a professional context. They had concerns that the feedback could add pressure to these inherently stressful situations. One participant stated mistrust: *"I would, if I were confronted with this [system], in an exam, in a job interview or a debate, where you have to show that you are socially capable but at the same time you can state your opinion, I would have a deep mistrust against this system, because I wouldn't know what it means (...)."* (A1P1). This indicates that transparency might be a crucial factor for professional systems.

Loss of Subtle Cues Anonymous feedback does not provide direct opportunities for receiving subtle cues such as mimics, gestures or tone from the feedback originator. This was experienced as limiting by our participants. *"I prefer to look the other people in the eye and pay attention to mimic and gestures, but less to such a computer thing."* (A3P2). Statements suggest that decoupling of body language from technologically mediated feedback is experienced as unnatural: *"You also get feedback on the natural way. If someone signals with glances or gestures that he found something unreasonable or interesting, that would* *be sufficient for me."* (A1P3); *"I find that the others should give their feedback during the discussion and not through the system. I would like that better. When he finds something bad he should react to it so that I get my feedback. I find that more pleasant."* (I4P1).

Loss of Granularity Intertwined with that is the limited granularity of the feedback. In our case, four different grades of feedback were offered. One participant stated: *"A system like that would be silly. That is only plus and minus, yes and no. It doesn't capture the state. In life there is normally not just yes or no."* (B4P1).

Positive Experiences

While similar concerns were raised in all conditions, group members reported about positive experiences only in the identifiable condition, although not all of the positive comments refer explicitly to the identifiability.

Overview Participants valued that the group mirror gives them the possibility to get a quick overview of the opinions in the group. Two group members stated: *"If one person dominates the discussion and the others do not even show their approval or disapproval through mimics, then you could get a better overview [with the group mirror]. How is the atmosphere, that would be fascinating to see (...)."* (I2P4); *"I found it pleasant, because you could incidentally notice what is going on."* (I2P3).

Everyone can Contribute Direct feedback through another person without the intermediate step of a group mirror is problematic in collaborative situations. The feedback interrupts the discussion and especially in large groups, it is difficult or even impossible that all participants give direct feedback to each other. In contrast, a group mirror enables participants to provide their feedback silently and does not interrupt the discussion. It also affords the opportunity that more group members can give feedback in large groups. Participants stated: *"I think it would be quite helpful to know, what the others think about it. I think it is helpful, when it is a large group and not everyone has the chance to contribute."* (I1P1); *"People that do not get a chance to contribute can also take part or state their opinion. The otherwise silent crowd that normally is overlooked."* (I4P4).

Engagement Furthermore, participants remarked that with the feedback through the identifiable group mirror they felt more engaged and that the group mirror could be helpful for stimulating discussions: *"You are more engaged. Actually, you have to listen."* (I4P3); *"It would definitely stimulate the discussion"* (I4P4).

Form Factor The form factor was the same in both conditions but was only positively mentioned in the interviews of the identifiable condition. Participants stated: *"I find the form factor quite charming"* (I2P4); *"Furthermore, the form factor is in such a way that you can also place it somewhere else"* (I2P2); *"Your [system] reacts extremely fast and good - interaction is fun"*. (I2P2).

Learning Participants stated that they reflected about feedback they provided and received. People set these positive remarks about how they worked with the group mirror less in context of leading a successful discussion but more in context of building well-formed arguments. One participant explained: *"In general, the feedback makes you think about what the feedback is like, what the others want to say. Not in terms of thinking about if I contributed something to the discussion but if my argument was valid and meaningful. I reflected a lot about what to do with the system."* (I2P2). Even negative feedback was valued in this context: *"If you get lots of negative feedback you have to think about the reasons. I think that is a good thing. There never is too much feedback, only too little. That's my opinion."* (I4P4).

Learning scenarios were mentioned as other appropriate use cases for our group mirror. Participants emphasized that feedback about argumentation to improve discussion skills is probably better supported by group mirrors than discussions with goals such as decision making or problem solving. Participants stated: *"It is much more important to know if my argumentation is good or not. The opinion people form is mostly predetermined."* (B1P3); *"Perhaps for politicians (...) when he should learn how to argue. In cases where you want participants to contribute balanced (...) - it would be interesting for schools."* (B2P4).

Comparison of Different Feedback Types

We explained the functionality of our group mirror to participants of the baseline. We showed groups of the identifiable conditions how the anonymous version works and groups of the anonymous version how identifiable feedback is provided. We then let them discuss their estimation of possible advantages and disadvantages of the two types.

Type (positive/negative) Participants supposed that identifiable feedback produces more positive feedback whereas anonymous feedback was assumed to generate more negative ratings. Two participants that had used the identifiable feedback discussed this aspect and reflected on a stronger hesitation for providing negative than positive feedback: *"I think you would get more negative [feedback], if it was anonymous"* (I2P2); *"Occasionally I indicated agreement although nothing terrific has been said, but I wouldn't indicate disagreement to something I do not agree to but that is not totally bad."* (I2P2). One reason is probably that people are afraid of the consequences of providing identifiable negative feedback. Several participants described a meeting with hierarchical structures as a situation in which they estimated identifiable feedback as problematic.

Familiarity An aspect that participants of the anonymous feedback condition thought could have an influence was the degree of familiarity between group members. In all groups, participants knew each other before the study. They stated that if people know each other well, anonymous feedback was superfluous but for group members that do not know each other, providing anonymous feedback might be more adequate. One group member said: *"We know each other for so long, I think we have the courage to state our opinion without thinking that the other person is offended too much."* (A3P2). However, one person (A1P3) stated that identifiable feedback would be better when the group members feel familiar, as the hesitation of giving negative feedback is inherently low in these situations.

Group Size Moreover, groups that had used the identifiable feedback perceived the size of the group as an influencing factor. They believed that anonymous feedback makes more sense for larger groups when it is not that important to see who provided the feedback but to get an idea of the overall estimation in that group. One group member explained: *"I think it depends on the size of the group. In a small group (...) when you do problem solving it makes sense to know who agrees and who disagrees. But in larger groups, yes [anonymous feedback makes sense]"* (I4P4).

Design Considerations

We discuss our results and derive considerations that can be taken into account when thinking about group mirror designs.

Tradeoff between Learning Effects and Acceptance Anonymous feedback, in terms of quantitative data, leads to a higher number of arguments and also to more well-formed arguments. That indicates that learning of building well-formed arguments was facilitated. One potential reason is that more feedback in general was given in the anonymous condition compared to the identifiable condition. Moreover, feedback was more negative which could mean that it also was more honest. The reason for this is probably that with identifiable feedback, more courage is needed to provide feedback.

Despite the increase in quantity of arguments and quality of argument structure, the anonymous feedback was perceived more negative than the identifiable feedback. Only one positive comment was made about the group mirror from a participant in the anonymous condition and more and sometimes also stronger negative remarks were made compared to the identifiable feedback. One participant of the anonymous condition expressed a potential *"deep mistrust"* (A1P1) against the group mirror in certain situations, which is probably attributable to the fact that people cannot ask the originator of the feedback, why he or she gave the feedback. This gap between the person and the feedback can then make people mistrust the group mirror. However, we want to emphasize that most of the negative comments were hypothetical and thereby more general concerns than actual experiences of the participants.

Although participants did not make many remarks explicitly about the type of feedback (anonymous or identifiable), an overall trend is visible in the answers of group members. Groups feel more comfortable with identifiable feedback. Participants of the identifiable condition reflected on several advantages and values of the group mirror, for example, the form factor, the advantage of including opinions from everyone in big groups or being more engaged and more thoughtful about the own arguments. One group even continued using the group mirror after the discussion during the interviews, showing that this group mirror is flexible to be used in other scenarios too. These advantages also exist with anonymous feedback, however, the negative aspects seem to outweigh these benefits as none of these aspects were mentioned in groups with anonymous feedback.

The just described aspects seem to imply a tradeoff between performance increase and acceptance of feedback. By providing the opportunity to give anonymous feedback, the inhibition level of providing especially negative feedback is lower so that the feedback is more valuable and can lead to a better learning effect. However, acceptance of the group mirror is lower and could impede with actual application of those systems.

Considering and Explaining the Use Case In the interviews, group members discussed appropriate use cases for our system. Although they only experienced discussions in a small group of four participants they assumed that the anonymous version of a group mirror was better suited for big groups while they preferred the identifiable version in the small group. They additionally took into account that they were familiar with the other group members. Most participants proposed the identifiable version for groups in which members know each other while the anonymous version was assessed more appropriate for groups in which group members do not know each other well.

The task the group mirror is used for is another important factor that has to be taken into account. Our participants assumed that our group mirror works better for a learning task compared to general discussions and meetings. The application of such a feedback system was perceived as problematic for using it in situations such as meetings or inherently stressful situations such as job interviews.

Moreover, it seems vital to clarify the realization and the purpose of the feedback. Results from the interviews reveal that participants are skeptical towards the use of group mirrors. This might derive from a lack of understanding of the reasons and benefits of providing technically supported feedback. This is also supported by the study Schiavo et al. (2014). They showed that subtle directives work better than overt directives, but only if people understand, how the feedback comes about and what it means. Therefore, a detailed briefing could improve the success of the group mirror. We assume that when explaining that our group mirror was about learning argumentation rules, the skeptical attitude that was sometimes expressed by our participants would have decreased.

Including Subtle Cues Group mirrors reduce feedback to special characteristics that seem most important in a certain context and normally cannot capture the whole complexity of feedback. In all conditions, participants remarked that they value seeing more information about the feedback than just yes or no, even more than verbal explanations. Mimics, gestures and tone seem to make feedback more valuable. Identifiable feedback meets this expectation more than anonymous feedback as the feedback provider is known and mimics and body language can be associated to the feedback.

It could therefore be beneficial to try to include subtler information that can get lost through a group mirror (such as mimics, body language and tone). This probably also implies a tradeoff between very simple and easily understandable feedback (e.g., using two colors to indicate approval and disapproval) and more fine-grained and complex feedback (e.g., using a graduation of colors).

Reducing Cognitive Load Especially for providing feedback a lot of information has to be processed to give valuable feedback. For learning tasks this might be appropriate but if the discussion itself prevails, reducing this cognitive load could make the tool less disruptive and the discussion more natural.

To reduce cognitive load in our scenario it could, for example, make sense to give the possibility to provide feedback about the existence of the individual parts of a well-formed argument instead of rating the quality of the whole argument. This would at the same time add valuable information to the feedback. Another idea could be to limit the possibility that everyone can give feedback all the time. It could make sense to choose one person randomly that has to give the feedback at a certain time. This would reduce cognitive load and disruption from the task and could also lead to a reduced risk of bullying as not all group members can give feedback simultaneously.

Guiding the Feedback Creation A system that uses peer feedback does not only need to take care of the presentation of the feedback but also of the generation of the feedback. The goal of the light cylinders was to show feedback about the quality of arguments. However, when participants were only asked to provide feedback about the quality, they often intermingled it with agreement and disagreement. To counteract this problem, we extended the system with the possibility to provide feedback about agreement and disagreement. Then, the differentiation worked well, though the smartphone app was more difficult to operate, as four buttons needed to be separated by the participants. In general, it might be valuable for future research to investigate these problems, on the one hand by analyzing the underlying processes that lead to these difficulties of differentiating between the concepts and on the other hand by exploring other solutions to that issue.

Summary and Conclusion

We introduced a system that enables group members to provide feedback to each other about agreement to arguments and argument structure. The system, composed of smartphones and light cylinders, can be used to provide anonymous feedback where the provider of the feedback is decoupled from the actual feedback and identifiable feedback where the speaker can see who gives what kind of feedback. A study comparing these types revealed the advantages of anonymous feedback. Results show significant differences compared to the identifiable version regarding the overall number of arguments and the quality of the argument structure. The increase of both of these factors indicates that learning to argue is facilitated. Qualitative data analysis however hints at possible problems with decoupling of feedback. We thereby added to the understanding of how technically provided feedback influences learning of basic argumentation rules and acceptance of feedback systems. We provided a number of considerations on usage scenarios for peer feedback group mirrors and on design of such systems.

An interesting question for further investigation is the difference between feedback directly from a person compared to technically mediated feedback. Our results indicate that the possibility to relate feedback to a person makes group members feel more comfortable. This comparison could give valuable insights in the aspects of verbal feedback that are important for the design of group mirrors. It could also be worth looking at different usage scenarios in more detail. Our participants concidered learning situations (learning how to argue) as appropriate, but other use cases, such as discussions in meetings, or arguing for learning a subject topic, may also be interesting and need further investigation.

A limitation of the study is that the identifiability was not the only independent variable. The placement of the light cylinders also changed. In the identifiable condition they were placed in front of participants while in the anonymous condition they were stacked on top of each other in the middle of the table. This makes it difficult to estimate the amount of influence the positioning of the cylinders had. Another study design would be possible in which the cylinders are stacked on top of each other in both conditions but the cylinders are assigned to individual group members. This setup will be evaluated in the following section.

7.2 Study II: Refinement and Further Evaluation

We conducted a second study with the prototype developed for the study described in the Section 7.1. We used the same prototype with a refined study setup. The study was conducted using a within-groups design instead of a between-groups design. To increase internal validity, the identifiability of the cylinders was realized by writing the names of the group members on the cylinders instead of positioning them in front of them. Our goal was to see, what effect this different way of making cylinders identifiable has on argumentation and the acceptance of the system.

7.2.1 Background and Motivation

The work described here builds on the work described in Section 7.1. The general background and motivation concerning the design of the prototype and the study are explained in detail there. We decided to conduct a second study with a refined study setup due to two circumstances. On the one hand, the study described in the previous section used a between-groups study design with 48 participants, nested in groups of four participants. This study design was chosen to reduce time commitment for participants and to reduce effects of fatigue. However, studies on group mirrors are often conducted using within-groups designs. This has the advantage that results are more expressive even with less participants. Thus, we decided to replicate the study using a within-groups design to verify the results.

Another observation that we made is that the positioning of the cylinders might have an influence on the results. Our initial goal was, among others, to investigate the differences between identifiable and anonymous feedback. In the identifiable condition, light cylinders were positioned in front of the participants. In the anonymous condition, all cylinders were stacked on top of each other in the middle of the table. This setup did not only change the identifiability of the feedback, but also the positioning (on the edges of the table or in the middle of the table). This might have influenced results, as feedback in the middle of the table might be easier to perceive at a glance than decentralized feedback. The refined design and the novel study setup are described in the following.

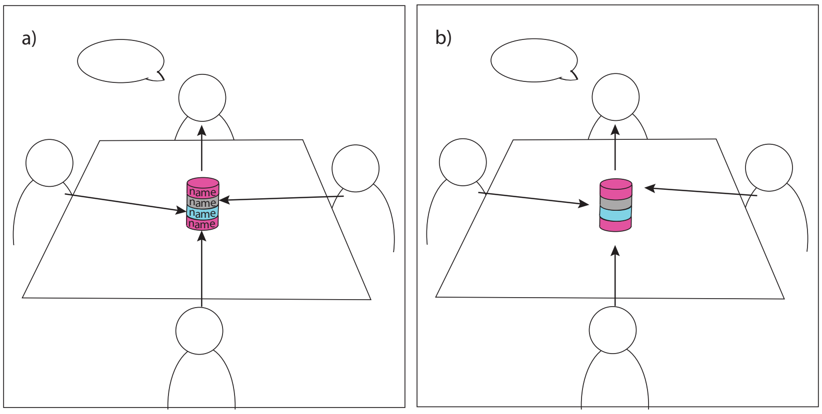


Figure 7.7: Study setup. In both conditions, light cylinders are stacked on top of each other in the middle of the table. a) Identifiable condition: names of all participants are additionally written on the cylinders b) Anonymous condition: feedback is sent to a randomly selected cylinder.

7.2.2 Concept and Design

Several possibilities exist to realize an anonymous and an identifiable condition without altering the positioning of the cylinders on the table. For instance, all cylinders can be positioned in the middle of the table in a circular arrangement. In the anonymous condition, the assignment would be randomized. In the identifiable condition, the cylinders could be aligned so that they point to the seating position of the respective group members. However, the assignment could be difficult to perceive for the whole group.

We decided to use a labeling of the cylinders to make the cylinders identifiable. Thus, we wrote the names of the participants on transparent adhesive tape that could be easily stuck on and removed from the cylinders. The name was written four times on each cylinder so that the names could be read from every seating position. The study setup is displayed in Figure 7.7.

7.2.3 Evaluation

Our main goal was to see, if results of a study using a within-groups design and a slightly different method of making cylinders identifiable validate or refute the results obtained from the study described in the previous section. That is why we conducted a study similar to the previous one and compared three conditions: a baseline without any support of a tool, a condition with an identifiable version of the system and a condition with an anonymous version.

Method

A within-groups design was chosen, due to the reasons described above. As in the first study, we compared three conditions: (1) groups without mediated feedback (baseline) (2) groups with anonymous feedback and (3) groups with identifiable feedback. Each group took part in each condition. We therefore needed three different topics. We counterbalanced conditions and topics using a Latin Square design.

Setup and Procedure

We took care to keep the setup and procedure as similar as possible to the first study. The study was conducted in the same room as the first study. Again, groups were built of four participants who sat around a rectangular table. In the beginning, a short introduction to the procedure of the study was given. First, the structure of well-formed arguments was explained. An example argument relying on the model of Weinberger et al. (2007) was given. Before the session that used the cylinders, we instructed the groups on how to use the group mirror. Before each condition, the topic was introduced. We chose three topics: A) *Is speed limit on freeways reasonable?*, B) *Should opening hours of shops be extended?* and C) *Should the voting age be reduced to 16 years?*. These topics are common in local political debates and the majority of people know about these issues. As in the first study, group members were asked to build well-formed arguments, composed of the parts *claim, ground* and *qualification* (Toulmin, 1958; Weinberger et al., 2007). Furthermore, a short test debate was held so that groups could get familiar with building arguments in a discussion. The topic of this test debate was *Legalization of marijuana*.

Then, each group discussed each topic for about seven to eight minutes. The detailed description of the three conditions can be found in Section 7.1.3. After each condition, participants filled in questionnaires with 5-point-Likert scales (1 = strongly disagree, 5 = strongly agree). Furthermore, we interviewed each participant individually after the study. All participants were thoroughly debriefed.

We collected quantitative as well as qualitative data. All sessions were audio and video recorded, discussions as well as interviews were fully transcribed and interaction data was logged. As described in Section 7.1.3, we analyzed the quantity of feedback, the quantity and quality of the arguments, the number of interruptions and the perception of the feedback.

Participants

We recruited 36 participants (17 female, average age 22 years, range: 18 to 35 years) who took part in the experiment in groups of four participants. All group members of each group already knew each other before the study and had not taken part in the first study with the cylinders. Most participants (35) were students of different subjects, one currently was enrolled in a Voluntary Year of Social Service. Participants were either reimbursed with a 10€ voucher from an online store or participated in the study as part of an obligation in their study program.

Quantitative Results

We evaluated data quantitatively and qualitatively. To compare the amount of feedback provided and the amount of arguments built between the three conditions, we used a linear

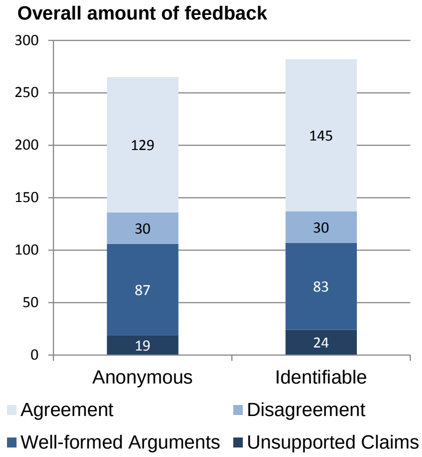


Figure 7.8: Results on the amount of feedback. The overall amount of feedback provided in the anonymous and the identifiable condition.

	Estimate	SE	DF	t-value	p-value
Intercept	8.16	1.34	60	6.11	
Anonymous	-0.47	0.73	60	-0.64	0.52
Subject A	-1.32	1.02	60	-1.30	0.20
Subject B	0.33	1.02	60	0.32	0.75

Table 7.4: Results of the linear mixed model for the amount of feedback. The identifiable condition serves as reference category.

mixed model with condition and topic as fixed effects and group as random intercept. REML (Restricted Maximum Likelihood) was used as a method of estimation. The model was fitted using the lme function from the package nlme (Pinheiro et al., 2015). This method was chosen because a repeated measures design with participants nested in different groups had been used in this study. The statistical software R was used for the analysis.

Quantity of Feedback The comparison between the amount of feedback provided in the identifiable and the anonymous conditions does not show significant differences. In the identifiable condition, slightly more feedback was provided (M = 7.83, SD = 4.49) than in the anonymous condition (M = 7.36, SD = 4.17) (see Figure 7.8). Table 7.4 shows the results of the linear mixed model.

Figure 7.8 also shows how much feedback of the different categories (agreement/disagreement and well-formed arguments/unsupported claims) participants provided. Feedback about well-formed arguments was provided more often in the identifiable condition. In the anonymous condition, 51.42% (absolute number: 145) of the feedback indicated agreement in comparison to 48.68% (129) indicating agreement in the identifiable condition. The percentage of well-formed arguments in contrast was slightly higher in the anonymous condition (32.83% (87)) compared to the identifiable condition (29.43% (83)). The percentage of disagreement (anonymous: 11.32% (30), identifiable: 10.64% (30)) and unsupported claims (anonymous: 7.17% (19), identifiable: 8.51% (24)) does not differ much between the two conditions.

Verbal Feedback We analyzed, which influence technologically mediated feedback has on verbal feedback. Hence, the amount of verbal feedback was counted. We developed a coding schema in which criteria and examples for positive and negative feedback were described. Two coders coded one discussion independently using the coding schema and the video recordings. Cohen's kappa showed substantial agreement between the two coders

	Estimate	SE	DF	t-value	p-value
Intercept	3.93	0.56	95	6.93	
Identifiable	-1.03	0.52	95	-0.64	0.05
Anonymous	-1.08	0.52	95	-1.30	0.04
Subject A	0.33	0.52	95	0.32	0.52
Subject B	0.38	0.52	95	0.32	0.46

Table 7.5: Results of the linear mixed model for the amount of verbal feedback. The baseline serves as reference category.

	Estimate	SE	DF	t-value	p-value
Intercept	0.83	0.24	95	3.52	
Anonymous	0.24	0.25	95	0.97	0.33
Identifiable	0.24	0.25	95	0.99	0.32
Subject A	0.09	0.73	95	0.35	0.52
Subject B	0.06	0.80	95	0.26	0.46

Table 7.6: Results of the linear mixed model for the quality of arguments. The baseline serves as reference category.

(κ = .89). We then compared the three conditions using the linear mixed model. Results show that more verbal feedback was provided in the baseline (M = 4.16, SD = 2.44) compared to the identifiable (M = 3.14, SD = 2.29) and the anonymous condition (M = 3.08, SD = 2.26). Results from the linear mixed model are summarized in Table 7.5. These results show that less verbal feedback was provided in presence of the system compared to the baseline.

Quality of Argumentation One of the two experts that had already rated the arguments in the first study also rated the arguments of the current study. A slightly adapted evaluation standard that normally is used in professional debate matches was again used. In the previous study, most arguments were rated as *unsupported claims*, as the evaluation standard of professional debates and the actual debate performance of inexperienced debaters differed quite a lot. Therefore, the evaluation standard was adapted to fit the current requirements better. Very good arguments are all arguments that contain a valid claim that is emphasized by a ground and a qualification (comparable to the *well-formed arguments* in the first study). Arguments that raise a valid claim, have a short explanation and lack a qualification are counted as *medium quality* arguments. Finally, *bad quality* arguments are arguments that have an explanation that is built on wrong assumptions. *Medium* and *bad quality* arguments are comparable to the category *unsupported claims* of the first study.

In the anonymous condition, the percentage of arguments of good quality was higher and the amount of bad arguments lower compared to the other conditions (see Figure 7.9). The percentage of good arguments was 26% (absolute number: 9) in the anonymous condition, 3% (1) in the identifiable condition and 15% (4) in the baseline. The percentage of bad arguments was 15% (5) in the anonymous condition, 21% (6) in the identifiable and 27% (7)

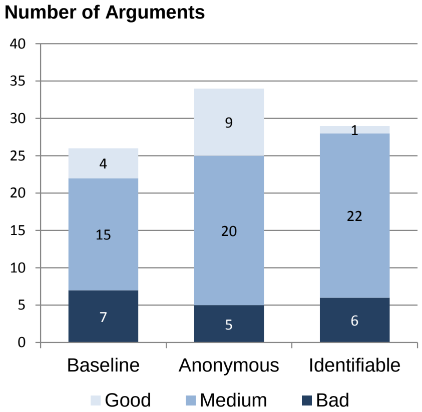


Figure 7.9: Results on the amount of arguments. The overall amount of arguments provided in the anonymous and the identifiable condition.

	Estimate	SE	DF	t-value	p-value
Intercept	0.68	0.21	95	3.27	
Anonymous	0.22	0.22	95	1.03	0.31
Identifiable	0.08	0.22	95	0.39	0.70
Subject A	0.19	0.22	95	0.90	0.37
Subject B	-0.05	0.22	95	-0.26	0.80

Table 7.7: Results of the linear mixed model for the amount of arguments. The identifiable condition serves as reference category.

in the baseline. The percentage of arguments of medium quality was highest in the identifiable condition (76% (22)), followed by the anonymous (59% (20)) and the baseline (58% (15)).

To conduct a statistical analysis on the quality of the argumentation, the three categories were assigned values from 1 (*bad quality argument*) to 3 (*good quality argument*). Then, the average quality of the arguments was calculated. We compared the average quality of the arguments between the three conditions using a linear mixed model. Results show that the medium quality of arguments is higher in the anonymous and the identifiable condition compared to the baseline. However, results from the linear mixed model did not show significant differences between the conditions (see Table 7.6).

Quantity of Arguments We compared the overall amount of arguments between the three conditions. The results of the linear mixed model did not reveal significant differences. Overall, most arguments (34) were built in the anonymous condition, second most (29) in the identifiable and the least in the baseline (26). Table 7.7 shows the results of the analysis. Figure 7.9 depicts the overall amount of arguments in the three conditions.

Preferences As in this study, participants got to know all three conditions, we asked them which condition they estimated as most satisfying. 12 participants (33%) stated that they did not notice a difference, 10 (28%) perceived the identifiable version as most satisfying and each 7 (19%) the anonymous and the baseline. This reveals the tendency that group members like the identifiable version more than the anonymous, however, differences seem weaker than in the first study.

Qualitative Results

In general, data gathered from the questionnaires and interviews point in a similar direction as the results from the first study, although the system was generally perceived more positively. The participants did not have that many concerns regarding the anonymous version compared to the first study and more positive remarks were made both regarding the anonymous and the identifiable condition. For instance, one participant stated: *"It was fun!"* (G2P4), another remarked: *"It is a good system, that you get [the feedback] immediately"* (G8P2). As we did not change the light cylinders themselves, the study setup might be one reason for this different perception. Due to the within-groups design, participants could compare the system-supported session to the unsupported session. This might have made the support through the system and the associated advantages more prevalent. One participant said that he or she missed the tool in the baseline: *"I found the last discussion [baseline] somehow... I would have liked to still have the feedback system. (...) I somehow missed it. It was actually nice to give feedback."* (G4P4)

In the interviews, we also asked participants, which version of the system they preferred and why. Most participants did not have a strong preference, several participants stated that it did not make much of a difference, but that they could see some advantages for one or the other version. From the 36 participants, 16 tended to prefer the anonymous version. The reason that was brought up most often was that feedback was *"more honest"* (9 times), followed by the reason that one has *"the courage to state your opinion"* (6 times). Other reasons were that the anonymous feedback is rather perceived as representing an overall mood of the group (2 times) and that reaction to the feedback are more directed to the whole group than only to individual group members (G1P2). One participant found the anonymous feedback *"more exciting"* than the identifiable feedback (G9P3).

The main argument in favor of the identifiable version was the speakers' ability to react to feedback directly. From nine participants that tended to like the identifiable version better, eight brought this up as a reason, *"because it just helps when you can react to the feedback"* (G7P2). One person said that the identifiable version was more comfortable because one did not need to enter the feedback secretly.

Discussion and Comparison of the Studies

One of the main goals of this study was to see, if we could replicate results of the first study (see Section 7.1.3) using a repeated measures study design and a slightly different way of making cylinders identifiable.

The overall amount of feedback that was provided did not differ significantly between the two conditions in either of the two studies. A general statement about the issues about achieving significant results in the studies presented in this thesis is provided in Section 9.2.5. However, in the first study, a tendency was visible that the identifiable condition produced less feedback. In the second study, this tendency is not visible. To understand the reasons for that is difficult, as the two studies cannot be compared directly due to their different study setups. One explanation might be that the different approach of making cylinders identifiable led to these different tendencies. While in the first study, it was easy to assign the feedback from the cylinders to the group members as they were placed in front of them, in the second study this was more difficult as one had to read the name to assign the cylinders to the group members. We assume that the more difficult it is to identify the originator of the feedback, the more feedback will be provided, which would be in line with findings by Nunamaker et al. (1996) or Dubrovsky et al. (1991), who argue that anonymity can lead to more and more equalized participation.

Another finding of the first study was that more arguments were built with the anonymous version. The second study only reveals a tendency of these results, but a statistical analysis does not show significant differences. There might be several reasons why the results only reveal a tendency. It could indicate that the increased amount of feedback in the identifiable condition in the second study led to an increase in the number of arguments. However, the influence of the anonymity on the number of ideas does not seem to be fully outweighed by the increase of feedback, as the tendency of more arguments in the anonymous conditions was still observable.

Further, the results of the first study show that argument quality was better in the anonymous condition. Again, a tendency supporting this result is also visible in the second study. Due to the different study setup and a slightly different evaluation metric, results are difficult to compare directly. Nevertheless, one explanation could be that in the anonymous condition of the first study, the highest percentage of negative feedback was provided. The increase in negative feedback could have made participants aware of missing parts of their arguments and could have nudged participants to improve their argumentation.

This negative feedback might also be accountable for the more negative attitude towards the system in the anonymous condition in the first study, as negative feedback can also lead to negative emotions (see e.g., Lazarus, 1991). The qualitative results show that in the first study, participants were especially doubtful when using the anonymous version. This effect is weaker in the second study. Another reason might be the repeated measures study setup. All participants used the system as an anonymous and an identifiable version. Knowing about both versions might have influenced the general perception of the tool.

Summary

We conducted a follow-up study to the study described in Section 7.1. This time, a within-groups design was used. Furthermore, we changed the way the cylinders were made identifiable to increase the internal validity of the study. For that purpose, the light cylinders were positioned in the middle of the table in both conditions. In the identifiable condition, names were written on the cylinders.

Results of the study show a tendency towards more arguments and higher quality in the anonymous condition, although differences are not statistically significant. While these results are not as strong as in the first study (in which differences in number and quality of arguments were statistically significant), they still corroborate the results. In contrast to the first study, group members did perceive the anonymous version as more positive. Moreover, we could observe that due to the possibility of providing feedback mediated through the system, less verbal feedback was provided in both the identifiable and the anonymous condition.

7.3 Chapter Summary

In this chapter, a prototype consisting of interactive light objects with the goal to support argumentation was presented. With the system, peers can provide feedback to other group members about agreement and quality of their argumentation. Two studies were conducted. The main goals of the first study were to evaluate the general feasibility of such a system to support people in learning how to discuss and to build proper arguments. Moreover, two versions were compared to each other and to a baseline: in one version, the feedback to the other group members was provided anonymously, in the other version, the feedback originator could be identified. The second study was conducted using the same prototype but using a slightly different study setup in order to increase internal validity. The main results of the two studies are as follows:

- Interactive light objects enabling peer feedback are an effective way of supporting co-located groups in discussions and can increase argumentation quality.
- Anonymous feedback tends to increase the amount of feedback in general and also the amount of negative feedback. At the same time, argumentation quality can be increased (see Figure 7.10, green arrow).
- When it is easy to identify the originator of the feedback, group members feel more comfortable (see Figure 7.10, yellow arrow).
- Verbal feedback decreases when groups get the possibility to give their feedback mediated through technology.

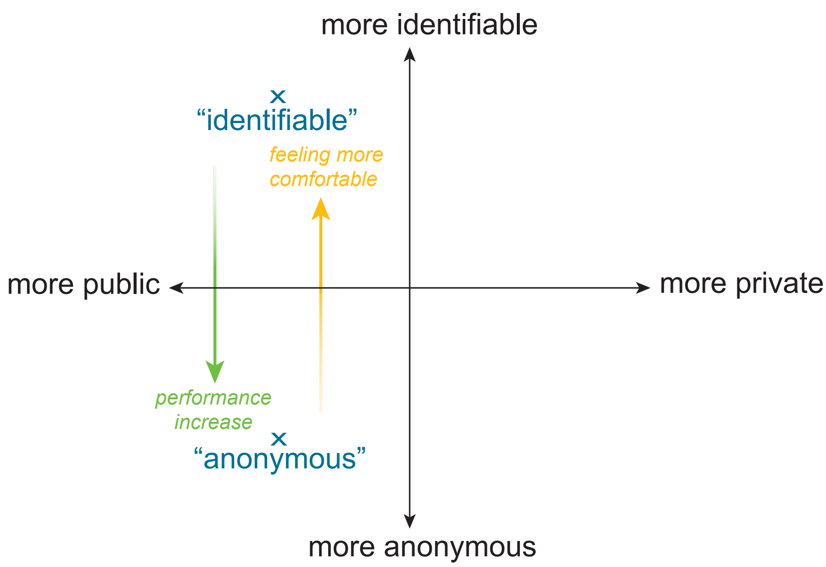


Figure 7.10: Overview of results. Results indicate that performance (i.e., number of arguments and argumentation quality) increases when feedback is given anonymously while people feel more comfortable when they know from whom the feedback comes.

- A number of design considerations are derived. For instance, when designing such peer feedback tools, a tradeoff between increasing learning effects and increasing the acceptance of the system might be necessary. A number of other factors also play a role (e.g., group size, familiarity of the group members).

8
Supporting Debates

After having investigated an interactive system supporting collaborative argumentation in the previous chapter, a more structured way of argumentation, namely debates, will be addressed in the following. We conducted two studies with a prototype designed to provide real-time feedback from a jury to the speakers during a debate. The first study was designed as a field study that was conducted as part of the two debates of a debate club. With this, we strove to obtain insights about how debaters who are actively involved in debate clubs accept feedback mediated by a group mirror and in which ways the feedback can affect debate performance. The second study builds upon that study and approaches the question on which level of detail information can be displayed to the speakers in real-time during a debate.

8.1 A Prototype for Direct Feedback during Debates

In this section, I will describe an interactive system designed to support real-time feedback during debates. After a short introduction discussing the background and motivation of this project, the design process is outlined, followed by the description of the field study that we conducted as part of two debates of a debate club.

This chapter is based on a collaborative work with Bernd Huber and a bachelor thesis (Rindlbacher, 2015). Part of it was published in the proceedings of a conference (Huber et al., 2014) with the co-authors Bernd Huber and Heinrich Hußmann. The detailed personal contribution statement can be found in the disclaimer.

8.1.1 Background and Motivation

Our goal was to develop a system to support debates, in this specific case the *British Parliamentary Style* debate (for an explanation of that debate style see Section 3.2.1), which is a debate style often used in debate clubs. These clubs aim, among other things, at teaching how to debate. Argumentative debating is an important skill in professional and educational contexts as well as in everybody's daily life. There is evidence that argumentation can increase engagement (Baker, 1999), lead to intellectual reflection (Mason and Santi, 1994) and support the acquisition of domain-specific knowledge (Baker, 2003; Kuhn and Goh, 2005). However, few people train and internalize the skill of debating and a good argumentation style. One of the reasons is that debating is hard to learn and requires a lot of practice and training (Kuhn et al., 1997).

The quality of an argument is highly related to its inherent structure, and therefore provides opportunities for technology to accelerate argumentation learning. Over the last 20 to 25 years, software tools to support and teach argumentation have been developed. Computer-supported collaboration scripts, for instance, support learners in constructing arguments by sequencing the argumentative activities (Baker, 2003; Stegmann et al., 2007). An overview on computer-supported argumentation can be found in Scheuer et al. (2010). A more detailed discussion of computer-supported argumentation is given in Section 3.2.2.

There is a number of studies that investigate feedback in oral presentations. The feedback tool HANS (Tam et al., 2013) showed that a vibrating wristband can help speakers with their time management and can facilitate a subtle interaction between a speaker and, in their specific case, the chair of a session of a conference. TALKZONES (Saket et al., 2014) uses mobile devices in order to allow speakers for more flexible time management. Commercial products such as *PowerPoint*[1] and *Keynote*[2] offer visual feedback in form of elapsed time and predefined speaker notes. PRESENTATION SENSEI (Kurihara et al., 2007) focuses on the preparation phase of a presentation. It aims at improving the delivery of the content by giving feedback about eye-contact to the audience, speaking rates and timing. The results of these studies show the potential of real-time feedback during presentations. An open question is, how qualitative feedback should be provided.

In debate clubs, a jury usually provides (qualitative) feedback to speakers of the *British Parliamentary Style* debate. However, this is done after the debate. As described in Section 3.2.1, the eight speakers have seven minutes each to present their arguments. Thus, one debate typically lasts about an hour, leading to a situation in which feedback from the jury is provided time-displaced to the actual situation to which the feedback is directed. The prototype presented in this section aims at helping beginners to integrate feedback immediately, leading to a shorter feedback-loop and faster learning (Bandura, 1977; Hattie and Timperley, 2007; Schunk, 1990).

[1] https://www.office.com/, last retrieved 04.07.2016

[2] http://www.apple.com/mac/keynote/, last retrieved 04.07.2016

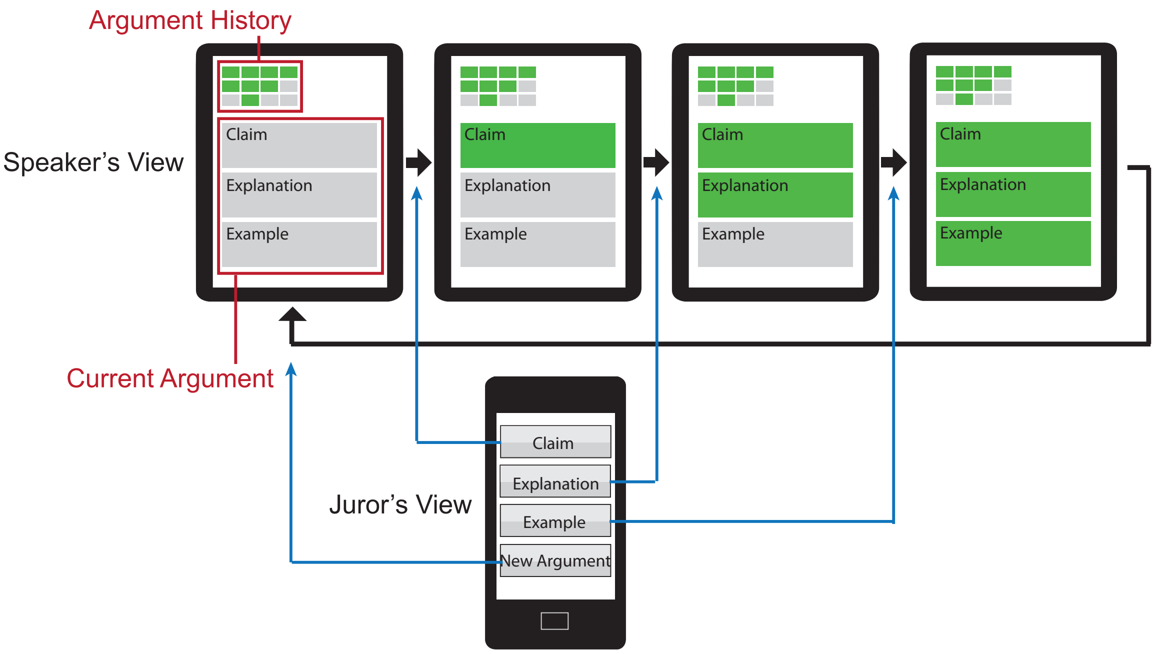


Figure 8.1: System design. The speaker sees feedback about past arguments and the current argument on a tablet computer. The jury provides feedback about the use of the argument parts using a smartphone.

8.1.2 Concept and Design

We decided to provide feedback about one of the most basic aspects that novices to debates need to learn: the structure of well-formed arguments. Arguments should be constructed of the three elements *claim*, *explanation* and *example*. This is one structure that is commonly used for segmenting arguments (a similar approach has, for instance, been proposed by Toulmin, 1958). We provide feedback about this structure by using a visualization composed of a combination of simple shapes and colors. This is intended to help speakers to perceive the information with a glance on the display during their speech.

Figure 8.1 shows, how the interface looks like for the speaker and the jury. The speaker interface shows the current argument and the history of arguments. The current argument consists of the three fields *claim*, *explanation* and *example* that are initially all gray. When the jury feels that the speaker has raised a claim, he or she presses the button with the label “Claim”. The *claim* field on the speaker's view then turns green. The same applies for the fields *explanation* and *example*. The buttons do not need to be pressed in that particular order and single fields can also be left gray when necessary. When the juror presses the button “New Argument”, the current argument moves to the top and is aligned right of the previous argument history to show, how many arguments have been built already and which arguments were built of all three parts or did miss certain parts. A new argument appears, meaning that all fields of the current argument on the speaker's view turn gray again.

Figure 8.2: Study setup. The feedback system is indicated with a red circle.

The system was implemented as a web-based interface using HTML5, CSS3, JavaScript and PHP. The two interfaces, one running on a tablet and one on a smartphone, communicate through a webserver.

8.1.3 Field Study

We conducted a study in course of the regular weekly meeting of a debate club. We decided to test the first prototype in a real world scenario. Our intention was to get feedback early in the design process both from people who are experts in debating, as well as from people who are less experienced but who are probably interested in learning how to debate as they did join a debate club of their own accord. Our main questions were whether speakers can understand and apply the information shown by the system and if or how much the feedback distracts from the actual task.

Method

We conducted the field study using a repeated measures design with two conditions: a baseline in which the speakers were not equipped with a tablet and a condition in which the feedback was displayed on a tablet.

Setup and Procedure

The study was conducted in a quiet room. The study took part on two dates with one week in between. Before each session, the experimenter gave a brief introduction to the study. The system was briefly introduced to the speakers and the juror. Then, the normal procedure of the debates began. In the beginning, the juror introduced the topic. On the first date, the topic was: *This house would send the prodigal son back*. On the second date the topic was: *This house would forbid pornography*. These are common debate topics and had been chosen by the debate club. Then, the speakers had 15 minutes to prepare their arguments.

During the debates, the feedback system was positioned at the speaker's desk for the sessions in which the speaker was supported by the system (see Figure 8.2). The juror controlled the system by using the interface on the smartphone, additionally to taking notes about the performance of the speaker as jurors usually do.

After the debates, questionnaires using 5-point Likert scales (1 = strongly disagree, 5 = strongly agree) were handed out. Furthermore, a semi-structured interview was held with the whole group after both debates. Both sessions were audio and video recorded. All input from the jury was logged with timestamp and the name of the button that was pressed.

Participants

In the study, 12 voluntary participants took part (2 female, average age 24, range: 18 to 42 years). All of them were members of the same debate club. Two of the participants can be seen as experts (took part in over 100 debates), three of them already had substantive experience in debating (20-60 debates), while seven were classified as novices (less than 10 debates). Four of the participants and one juror took part in both sessions. The other eight participants (four in each session) were different people in both sessions. We gathered qualitative feedback about their experience with the system from the other participants that had used the system only once.

Results

We evaluated the study by measuring the gaze direction of the speakers using the video recordings and by evaluating the reported self-efficacy using the questionnaires (Bandura, 1977; Schunk, 1990). Qualitative results were gathered from the interviews.

Acceptance

In general, most participants regarded the feedback system as a valuable support due to the immediate feedback. Participants stated that the *"real-time feedback allows for direct improvement."* (P4), or that *"when all three parts are green, I receive an extra push, like a reward system"* (P5).

We asked participants if they thought that the separation into *claim*, *explanation* and *example* was helpful for them. Figure 8.3 shows the answers from the questionnaires. One participant (10%) strongly agreed and five participants (50%) agreed that this was the case. Two of these participants were experts in debates (more than 100 debates) and four were novices (less than 10 debates). Three participants (30%) disagreed and one participant (10%) strongly disagreed. Two of these participants were novices. Both stated in the comments that they either did not get much feedback or did not pay attention to the feedback. The other two had substantive experience in debating (20-60 debates).

These results indicate that especially experts as well as novices who are open to being supported by the system perceive the feedback about the arguments structure as helpful. The one novice speaker who did not receive much feedback did not structure arguments in the way the system suggests, which leads to a situation in which the feedback is less valuable.

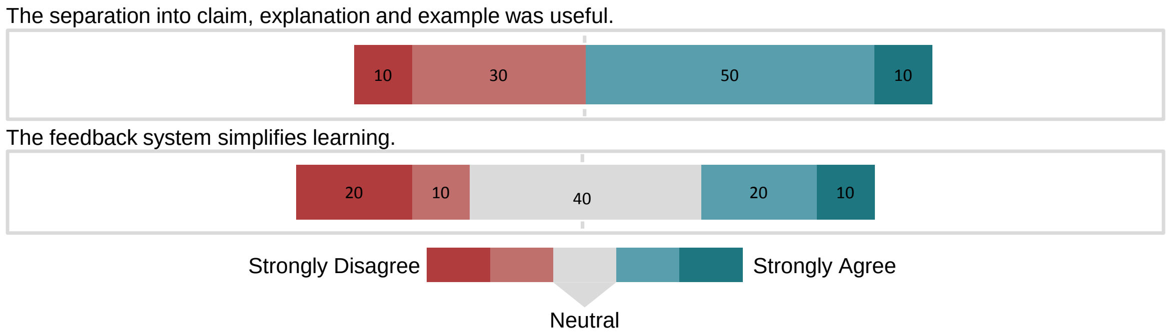


Figure 8.3: Results of the questionnaires. Answers to the questionnaires about the structure of arguments and how participants perceived that learning was supported. Numbers indicate the percentage of participants who answered with that score on the 5-Point Likert scale.

The speaker who did not pay attention to the feedback did not believe that he or she could have benefited from the feedback. In the comments, he or she stated: *"I don't believe that I could have changed my speech in real time."* (P11)

We, moreover, asked participants whether they thought that the feedback could simplify learning. Four participants (40%) were neutral regarding this aspect, one (10%) strongly agreed, two (20%) agreed, one disagreed (10%) and two (20%) strongly disagreed. This divergent opinions are also reflected in the interviews after the debate. One participants stated that *"when you are a good debater you don't need it and if you are a poor debater it doesn't help you (...) but will confuse you."* (P5). Another participant on the contrary thought that it is *"super for beginners, because it is the basis of debating to effectively convey your arguments"* (P1).

Distraction To evaluate how distracted participants were due to the system, we counted the number of gazes to the system and compared it to the number of gazes to the desk or to the notes in the baseline. All participants used hand written notes that were lying on the table or which participants held in their hands. The average amount of gazes to the desk was 20 without the system and 21 with the system. This shows that the participants did not look at the system more often then they would look at their notes. Additionally, gazes mostly endured less than one second. However, the system was also perceived as *"sometimes distracting"* (P3) and *"too inflexible"* (P4).

Estimation of the Jurors In the interviews, we asked the jurors about their experiences with the system. They could take notes, follow the debate and provide feedback through the system without much problems. However, sometimes it was difficult for them to decide which feedback to provide in case a speaker did not follow the given structure at all.

Suggestions for Improvement Several proposals were made by the participants, both by experts as well as by novices and the jurors. Especially the jurors would have liked to have more flexibility in providing feedback. In this prototype, an argument part is either absent or present. However, it is often the case that speakers build arguments of the parts *claim*, *explanation* and *example*, but do not elaborate much on the details. The jurors therefore recommended to introduce a more fine-grained feedback, for instance, by using colors: yellow could indicate that the argument part has been mentioned, but still needs more elaboration to receive a green rating.

Furthermore, several participants proposed to include more feedback than only the one on the structure of an argument. The juror reported about a debate he had attended in which one person gave the speakers feedback by raising pictures showing symbols about the structure, loudness, pace etc. Several participants shared the estimation that more information than just the structure of an argument can effectively be perceived and included during a debate.

Another suggestion for improvement was to include a display of the elapsed or remaining time into the system. In this study, debaters brought their own timer and positioned it next to the system.

Summary and Discussion

In this section, I described a prototype designed for supporting debaters in structuring their arguments. We take advantage of the fact that in debates as practiced in debate clubs, jury members are present who rate the performance of the speakers. Our system uses the ratings of the jury to provide real-time feedback to the speakers with the goal to allow debaters to improve their speech immediately.

We conducted a field study in the course of the debates of a debate club to get some preliminary feedback on our prototype both from experts and novices who are interested in debating. Results indicate that participants generally liked the feedback. Especially experts and novices valued the feedback. One assumption why the two experts of our study valued the feedback system might be that they see the use case of training novices with such a tool. Novices who believed that structuring arguments in the three parts (*claim*, *explanation* and *example*) is valuable and who thought that the feedback could effectively support them in learning how to debate, also valued to system.

A number of areas for improvement were recommended, such as showing more information in a more flexible way. In the next section, I will present a system that is based on these suggestions and describe a study that investigates how much information speakers actually can process during a debate.

8.2 A Comparison of Different Complexity Levels

Based on the results of the field study described in the previous chapter, we investigated, how much information speakers can effectively process in real-time during a debate. Therefore, we built another prototype and compared three conditions in a study: a baseline, in which the system only shows the remaining time, a condition in which the tool shows similar feedback compared to the prototype that we used in the field study (i.e., structure of the arguments and history of past arguments) and a condition that additionally shows text cues.

8.2.1 Background and Motivation

As described in Section 8.1.1, learning how to debate and to build arguments is an important skill. In debate clubs, a jury rates the performance of the speakers and provides feedback after the debate. Our main goal was to make use of the possibility to provide real-time feedback from the jury to the speakers in order to shorten this feedback-loop. In a study, we were interested to understand the influence of content-related real-time feedback on debating performance and learning how to debate.

In the previous section, a field study was discussed using a first prototype that allows a subtle teacher-student dialog between jury and speaker. First results indicate that speakers can effectively include the feedback in their speech in real-time. We collected recommendations for improvement both from experts as well as novices. We adapted our prototype based on these insights and conducted a controlled laboratory study to investigate the effects of more or less complex feedback.

8.2.2 Concept and Design

Based on the results of the field study (see Section 8.1), the results of a brainstorming session with three participants (two had attended a course on debating before) and an interview with an expert with longstanding experience in debating and membership in a debate jury, we designed three versions of the system.

Baseline The interface shows the remaining time of the designated seven minutes (see figure 8.4, left). The jury interface allows for counting claims, explanations and examples for the subsequent analysis, but this information is not displayed on the speaker's device.

Structure Feedback There are several approaches to structure arguments (see e.g., Toulmin, 1958). In debates, the most common structure consists of the three elements *claim*, *explanation* and *example*, which we used in our system. As participants from our field study missed the possibility to get more detailed feedback, we extended the system with more fine-grained feedback on argumentation quality. While in the earlier prototype, the jury could indicate if a certain part of an argument (*claim*, *explanation*, *example*) was present, now the jury is able to indicate how well the specific part is covered by choosing a color (see figure 8.4, center).

In the beginning, all fields of an argument are grey. The color can change several times, an example can, for instance, be rated as "red" in the beginning and then be changed to yellow and green when the speaker continues his or her explanation. When a new argument begins, the jury will click on "New Argument" on their interface. On the speaker's tablet, the current argument is added to the history on the right side of the interface. A new argument with gray fields appears. Red means that this part is mentioned, but not sufficiently elaborated, green symbolizes that the part is explained very well and is understandable in all its facets. Yellow indicates a state in between. If a part is not mentioned at all (e.g., example missing) the field remains gray.

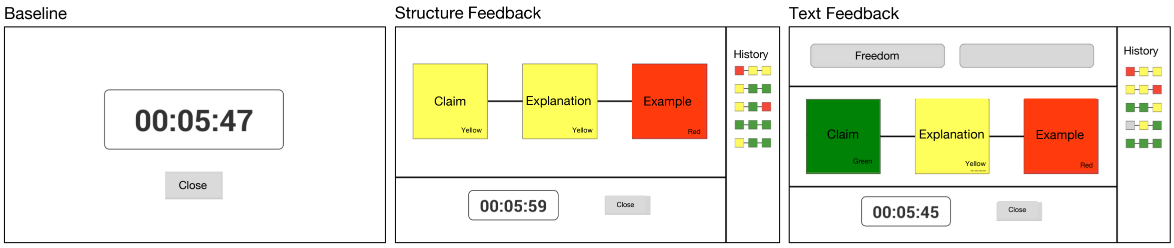


Figure 8.4: System design. Left: In the baseline, the interface shows the remaining time. Center: In the condition with structure feedback, feedback about the quality of the three parts of an argument is additionally displayed with the history of all arguments on the side. Right: Additional fields show textual hints (one for predefined text cues, one allowing free text entry).

Text Feedback While the basic functionality is equal to the structural feedback, we added the possibility to provide text cues to the speaker (see figure 8.4, right). This can be accomplished in two ways. The jury can choose predefined key words that are often used during debates to question an argument: “why", “whereby", “how” and “so what". The latter indicates that the broader implications of an argument are missing. In addition, there is a field to enter own keywords.

The system is implemented for tablets as a server-client architecture in Java using the Android SDK.

8.2.3 Evaluation

We conducted a study to investigate the influence of qualitative real-time feedback on debating. We were particularly interested in evaluating whether feedback about the argument structure can improve argumentation and if textual cues can be useful for advanced debaters, while structural feedback can be useful for beginners. We used different evaluation methods to figure out which are most appropriate for evaluating real-time feedback systems.

Method

The study was designed as a repeated measures experiment. Each participant accomplished three debates with the different feedback systems as independent variable. We therefore had to choose three different topics. The three conditions and topics were counterbalanced using a Latin square design. Each combination occurs once in each round (first, second or third debate of one participant).

Setup and Procedure

The study room was equipped with a table for the speaker and a table and chair for the juror. The table was positioned so that the speaker had a good view on the tablet (see Figure

Figure 8.5: Study setup. A speaker supported by the system with additional handwritten notes.

8.5) The jury was sitting at a table opposite to the speaker, as this is the standard setting of professional debates.

After a general introduction to the study procedure, we provided further information on debates. In particular, the importance of building proper arguments was explained. Therefore, an example of an argument consisting of *claim*, *explanation* and *example* was given. Before each debate, the topic was introduced. We decided upon three common debate topics: (1) *Pro vegetarianism* (2) *Pro school uniforms* and (3) *Pro Legalization of drugs* (International Debate Education Association, 2004; Sather, 1999).

Each participant completed three debates. For each debate, the participant had seven minutes time to think about arguments and take notes, which could be used during the debate. The experimenter then explained the feedback system. According to the British Parliament debate[3] (see also: International Debate Education Association, 2004), which is a common debate style in debate clubs, participants had up to seven minutes to state their arguments.

After each session, participants filled in pen-and-paper questionnaires with 5-point Likert scales (1 = strongly disagree, 5 = strongly agree) and a 5-minutes semi-structured interview was held. In the end, a final questionnaire including questions about demographics was handed out. Overall, the study lasted about one hour. Participants were thoroughly debriefed after the study.

The study was audio and video recorded. One camera filmed the participant from the front and one filmed the tablet. A microphone was positioned in front of the participant. In addition, we measured the heart rate with a *Polar* heart rate sensor[4], and affect in voice (i.e., emotion) with an automatic speech affect recognition toolkit *openSMILE* (Eyben et al., 2013).

3 http://idebate.org/about/debate/formats, last retrieved 06.07.2016

4 http://www.polar.com/en/products/accessories/H7_heart_rate_sensor, last retrieved 06.07.2016

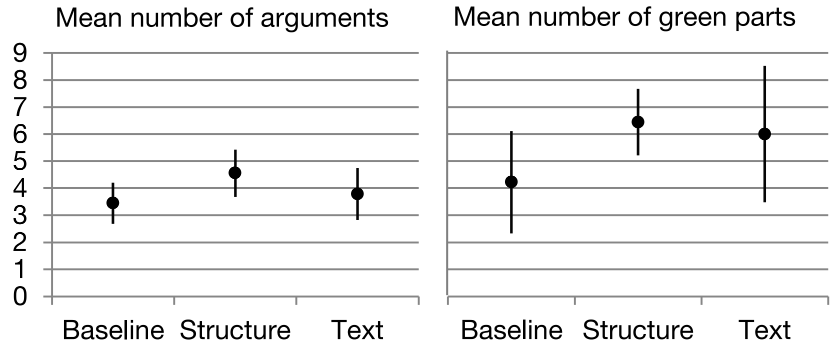


Figure 8.6: Results on the number of arguments and high-quality arguments. Diagrams showing means and 95% confidence intervals. Left: mean number of arguments for the three conditions. Right: mean number of parts of high quality ("green parts").

Participants

We recruited nine participants (4 female; average age 26, range: 23 to 28 years). Seven were students, one was a consultant and one did not make a declaration. None of the participants had taken part in a professional debate before. In addition, we invited an expert in debating with longstanding experience in debating juries to control the interface and provide the feedback to the participants. This was the same person that took part in the expert interview, which informed the system design. All participants received a voucher from a well-known online store.

Results

We evaluated the system using the following measures: the argument quality by using the ratings of the jury, the participants' experience with the system using pen-and-paper questionnaires and semi-structured interviews, physiological reactions using measurements of the heart rate and affect in voice with a speech affect recognition toolkit. To gain experiences on an alternative evaluation method, we report effect sizes with 95% confidence intervals (*CI*), an approach that is used with increasing frequency, as concerns with null hypothesis significance testing for reporting and interpreting results are growing in various fields (Cumming, 2013; Dragicevic, 2015).

Number of Arguments We counted each argument that consists of at least two parts as an argument (e.g., *claim* and *explanation*, *example* missing). Figure 8.6 (left) shows that more arguments were introduced in the condition with feedback about argument structure (M = 4.55, 95% CI = [3.68,5.43]) compared to the condition with additional textual cues (M = 3.77, 95% CI = [2.81,4.74]) and compared to the baseline (M = 3.44, 95% CI = [2.68,4.21]). That indicates that especially feedback about the argument structure stimulates speakers to produce more arguments. In the next paragraph, we will discuss, if these are also of higher quality than in the other conditions.

Affect	arousal	valence
Highest motivation with textual (# 3)	0.61	0.74
Highest motivation with structural (# 4)	0.10	0.42

Table 8.1: Results on arousal and valence. Values represent the percentage of high arousal and positive valence, measured for participants that were most motivated with structure feedback and participants that were most motivated with text feedback.

Quality of Arguments As the influence of the feedback system on the quality of the arguments was one of the main outcomes investigated, we looked at this aspect in more detail. Participants built the most argument parts that were of high quality (rated as "green") - for example, an explanation that was very detailed and elaborated - in the structure condition (M = 6.44, 95% CI = [5.21,7.68]) (see Figure 8.6, right). In the baseline, fewer argument parts were of high quality (M = 4.22, 95% CI = [2.33,6.11]). The mean of the condition with textual cues (M = 6.00, 95% CI = [3.48,8.52]) lies only slightly inside of the confidence interval of the baseline. The variance of the number of high quality parts is relatively high in this condition. This could indicate that textual cues provide a major support for some speakers (e.g., more experienced speakers) whereas others (e.g., less experienced speakers) may profit most from structural feedback. At the same time, the number of low quality arguments rated with "red" is less frequent in the text feedback condition (M = 0.11, 95% CI = [-0.09,0.32]) compared to the baseline (M = 0.78, 95% CI = [0.18,1.38]) and the structure feedback condition (M = 0.78, 95% CI = [0.11,1.45]).

Affect in Speech Because stress may have an important influence on debate learning, we quantified the affect of speech for each debate. We used the *openSMILE* toolkit to quantify arousal and valence of each debate (Eyben et al., 2016, 2013). Speaker affect was measured on the basis of 10 second frames and was averaged over all users who self-reported being more motivated to use structural feedback (4 participants x 3 debates) and all users who self-reported being more motivated to use textual feedback (3 participants x 3 debates), as seen in Table 8.1. In our study, speakers that self-reported to be more motivated with textual cues had higher arousal and valence values. This is an indication for higher speaking skills (Hirschberg and Rosenberg, 2005). Speakers with lower affect values were more motivated with structural feedback.

Speaking Time In professional debate, speakers have seven minutes to present their arguments. This may be short for a professional debater, however for novices it is difficult to fill this time. Accordingly, none of our participants used the full seven minutes. Participants without feedback probably did not know how to improve an argument further. However, with feedback (especially with text feedback), they received feedback about which aspects of an argument they could elaborate more. This can be seen in the results, which indicate that with a feedback system, especially with textual cues, participants are able to use more of their time (Text: M = 202.55, 95 % CI = [157.85, 247.26], Structure: M = 186.77, 95 % CI = [157.21, 216.34], Baseline: M = 166.88, 95 % CI = [135.51, 198.26]) (see Figure 8.7, left).

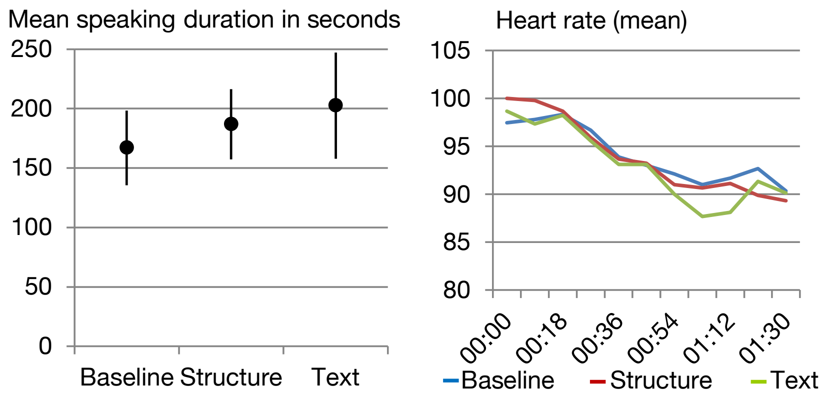


Figure 8.7: Results on speaking times and heart rate. Left: Results indicate that on average, participants spoke longer with text feedback compared to the other conditions. Right: The heart rate dropped during debating, however similarly in all conditions.

Heart Rate We hypothesized that physiological measurements can give us further insights in the effects of different types of feedback on argumentation. We measured the heart rate of participants every 9 seconds. For each feedback type, the heart rate was averaged over all speakers. However, results do not reveal strong differences between the three conditions (Text: $M = 90.77$, 95 % $CI = [82, 99.53]$, Structure: $M = 91.35$, 95 % $CI = [82.65, 100.06]$ Baseline: $M = 92.75$, 95 % $CI = [83.24, 102.27]$). Typically, the heart rate was higher during the debate phases compared to the preparation phases. In the debate phases, the heart rate of most participants decreased slightly over time, however similarly in all conditions (see Figure 8.7, right).

Reaction to Textual Cues We evaluated if participants noticed textual cues and if they included them in their argumentation. Overall, the juror gave 18 text cues during the study. Participants noticed twelve of these suggestions. From these, ten were incorporated by the speaker during the debate. Participants who seemed more experienced in giving presentations could incorporate these text cues successfully and could in this way strengthen their arguments. However, participants that seemed more inexperienced got confused and interrupted their argument. One participant, for instance, asked the juror for an explanation of the text hint.

Qualitative Analysis In the questionnaires, we asked participants to rate their experiences with the systems on a 5-point Likert scale (1: strongly disagree, 5: strongly agree). We asked them to what extent they paid attention to the argument structure (see Figure 8.8, top). Results show that they did this especially with text feedback, where all participants agreed (2 participants) or strongly agreed (7 p.) and with structure feedback (3 p. agreed, 6 p. strongly agreed). In the baseline, one participant did not agree, one was neutral, the rest agreed (3 p.) or strongly agreed (4 p.). We also asked how much participants felt interrupted (see Figure 8.8, middle). Results show that people feel more interrupted with more feedback: in the baseline all participants disagreed (5 p.) or strongly disagreed (4 p.), with text feedback five participants agreed (2 p.) or strongly agreed (3 p.). While the structural feedback was

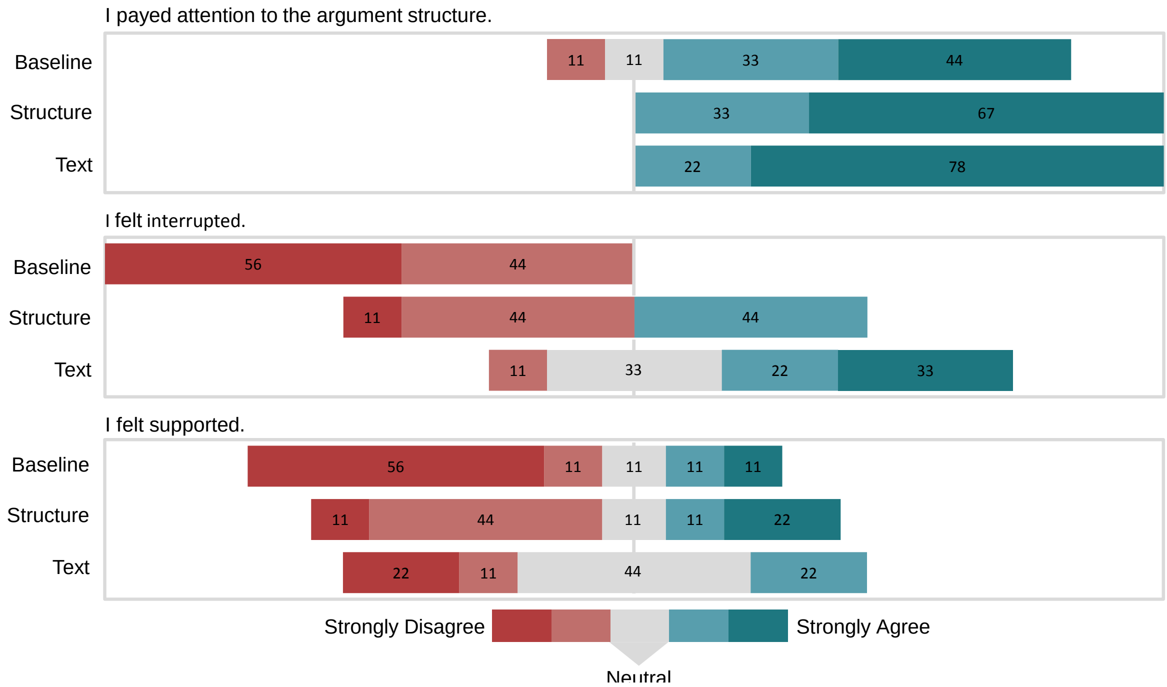


Figure 8.8: Results of the questionnaires. Answers to the questions about estimations of how much people payed attention to the argument structure, about feeling interrupted and feeling supported. Numbers indicate the percentage of participants who answered with that score on the 5-Point Likert scale. Numbers are rounded and thus might not add up to 100% exactly.

perceived as interrupting by four participants, five did not agree (4 p.) or strongly disagreed (1 p.). At the same time, participants felt slightly more supported with feedback (see Figure 8.8, bottom) (Structure: strongly agree: 2 p., agree: 1 p., neutral: 1 p., disagree: 4 p., strongly disagree: 1 p.; Text: strongly agree: 0 p., agree: 2 p., neutral: 4 p., disagree: 1 p., strongly disagree: 2 p.) compared to the baseline (strongly agree: 1 p., agree: 1 p., neutral: 1 p., disagree: 1 p., strongly disagree: 5 p.).

In the final questionnaires, we asked participants with which system they thought they performed best. Six participants named the system with structural feedback, three mentioned the baseline. Reasons that were brought up in the interview for the baseline were mostly that people felt overwhelmed or distracted by the other systems: "*The other systems distracted me too much and I had to look down to react to them. With the clock, I look at it briefly and know how much time is left, that's enough*" (P1). Reasons for the structural feedback were that "*it motivates that you see the boxes and that you want to have green boxes, but it doesn't distract too much*" (P3). Another participant stated: "*I didn't feel insecure. When it was yellow, it was also ok, but with the other one [text] you question yourself more*" (P6).

However, when asked with which feedback they would achieve the best learning success, eight voted for the text feedback, only one voted for the structure feedback. Reasons named for the text feedback were, for instance, that "*you notice where the expert or the audience* *did not understand something or where you left questions unanswered*" (P7). Another reason was that you "*(...) get counter questions immediately and can learn that there could possibly come this or that counterargument*" (P4).

Summary and Discussion

This section discussed an approach to provide real-time feedback during argumentative debate. We compared three feedback versions. One serves as a baseline and shows the remaining time, one additionally displays feedback about the argument structure and the third one displays the remaining time, the argument structure and additional textual feedback. Next to evaluating the quality of the arguments, we analyzed the participants' physical state through heart rate, as well as their expressivity by analyzing the speakers' affective characteristics. The results reveal the potential of real-time feedback during debates.

In particular, results indicate that more arguments and at the same time arguments of higher quality were built when speakers were supported with the structural feedback. Textual feedback seems to support more experienced speakers better than novices. This interpretation is supported both by the ratings of the experts as well as through the values obtained from the evaluation of the affect in speech.

Qualitative results indicate that debaters feel supported and pay attention to the argument structure more when supported with feedback. However, the more feedback is displayed, the more participants feel interrupted. This strengthens the use case that we see for such real-time feedback: in the phase of learning rather than during regular debates.

A limitation of that approach is its constrained way of dealing with argument structures. In the future, our approach could be combined with machine intelligence for a more adaptive argumentation model. One example for such a scenario would be to automatically analyze the spoken content, and then use an argument structure that fits the current argument's content. (i.e., switch between textual and structural feedback). Furthermore, our system may be used to provide scalable feedback for online lecturers by their audience.

8.3 Chapter Summary

In this chapter, an approach was discussed that had the goal to support speakers of debates in real-time with qualitative feedback provided through a jury, mediated through technology. A first prototype was evaluated in a field study to gather feedback from both novices and experts in debating. This prototype visualized how the speakers structure their arguments (*claim*, *explanation* and *example*). Based on these results, we designed and evaluated a prototype that came in three versions: a baseline showing the remaining time, a version showing feedback about the argument structure as in the first study and a version that additionally included the possibility to show text cues. The main results of these two studies are as follows:

- Real-time feedback about qualitative aspects from a teacher (such as a jury member) to a speaker is an adequate opportunity to help learners improve their argumentation immediately. The juror is capable of providing this feedback during the speech. Moreover, it is possible for the speakers to integrate the feedback from the jury directly.
- The feedback about the structure of an argument is helpful both for experts as well as for debate novices.
- For people who are less experienced with debating, the feedback about the argument structure can increase the number of arguments as well as the quality of arguments immediately.
- However, feedback about the argument structure seems not to be valuable for people who are not willing to or not capable of following a given argumentation structure. Similarly, feedback seems less valuable for people who think that it is not possible to incorporate the feedback in real-time.
- Speakers with higher or lower experiences in debating differ in how much feedback they can process. For novices, feedback about the argument structure seems to be more effective than only displaying the remaining time. Additional textual cues however only show benefits for more experienced speakers.

V

REFLECTING ON GROUP MIRRORS

9
Conclusion and Outlook

The main objective of this thesis was to get a better understanding of the influence of *group mirrors* on co-located collaboration. To approach this topic, I first presented a design space for group mirrors. Systems that had been developed in related research were classified according to this design space. Based on this systematic analysis of existing work on group mirrors, we developed seven different prototypes, which were evaluated in ten distinct studies. We applied two use cases. These are collaborative creativity and argumentation. In the next sections, I will classify the group mirrors presented in this thesis according to the design space. I will then assemble the different results of the studies and summarize the main insights. Moreover, I will discuss limitations of these studies and suggest areas for future research.

9.1 Summary and Classification

Figure 9.1 shows the projects described in this thesis classified according to the design space. A detailed explanation of this visualization can be found in Section 4.4. The classification of related work conducted in that section is also visible in this figure in form of light gray lines. The different projects of this work accompanied with the chapter in which they are explained are indicated at the top using different colors (e.g., *Groupgarden* (5.1), in pink). The different characteristics of the design space are displayed on the x-axis. The aspects that we did not incorporate in this work are displayed in light gray, aspects that were used at least in one of the various projects are highlighted in light green. For the aspect of "space" for instance, we only considered group mirrors in *co-located* and not in *remote* scenarios. Characteristics that we compared in a study are displayed in the color that is assigned to that project. For instance, with the system GROUPGARDEN, we compared two public display settings to each other, namely a visualization on table or wall. Thus, the aspects *public (table)* and *public (wall)* of the category "placement and privacy" are colored in pink.

What can also be read from this classification is that the goal of this thesis was not to reinvent the concept of group mirrors or to develop entirely novel and innovative group mirror

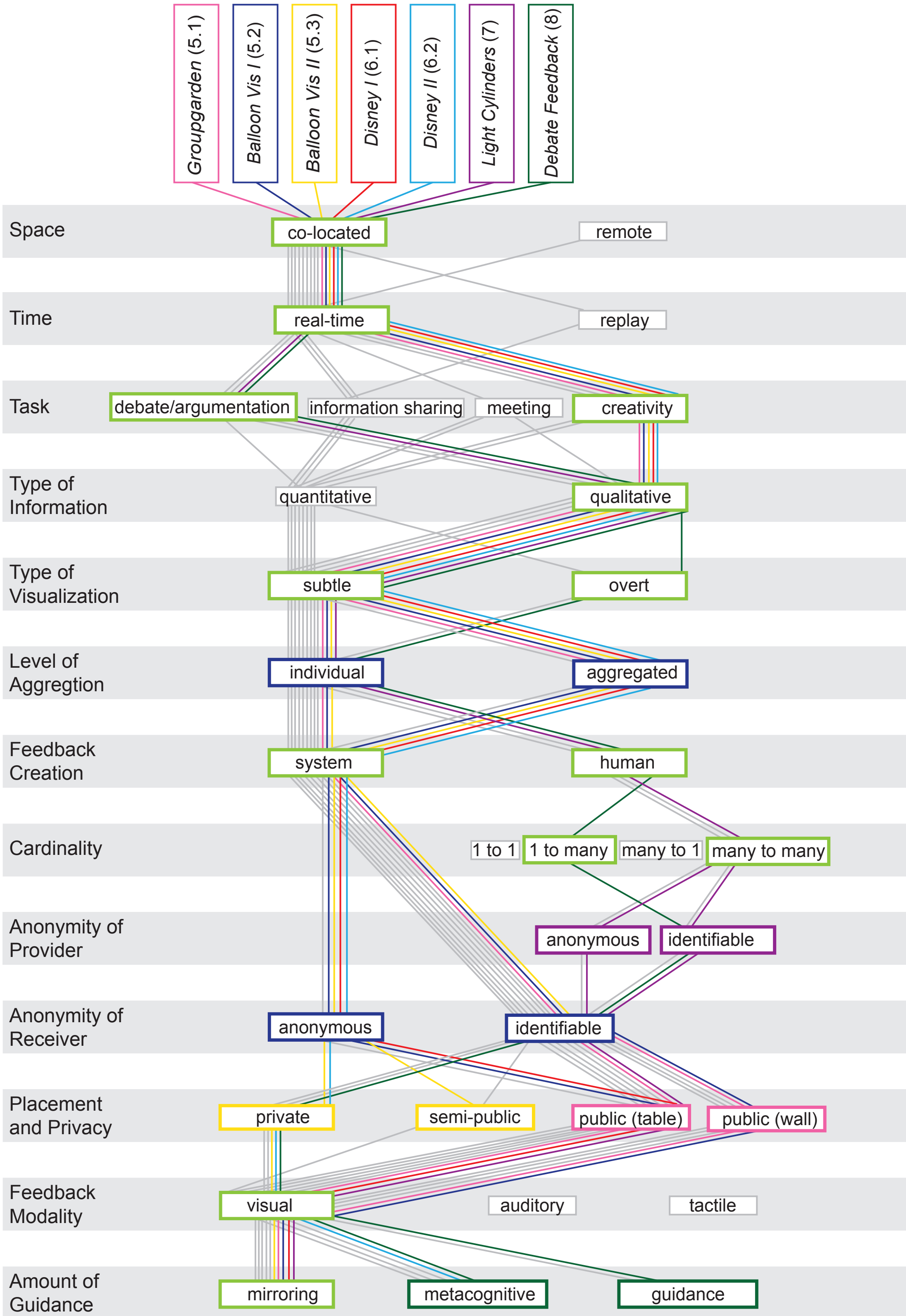


Figure 9.1: Classification of the projects. The projects presented in this thesis are classified according to the design space.

systems. In that case, hitherto mainly disregarded concepts, such as group mirrors using *auditory* or *tactile* feedback, would have been interesting areas for examining. Instead, we decided to build on existing research and to try to understand the effect of group mirrors better by investigating some selected characteristics. Thus, we made some initial design choices that align with the concept of most existing group mirrors.

Space and Time All our systems are designed for *co-located* collaborative scenarios in which feedback is displayed in *real-time*.

Task We chose two exemplary tasks (*debate/argumentation* and *creativity*) that represent the extremes of a continuum of tasks ranging from more structured to more open-ended. This was done to cover a wide range of possible use cases.

Type of Information The type of information was one of the categories in which our group mirrors differ from the majority of existing group mirrors. While existing systems mostly mirror *quantitative* information (such as speaking times), we decided to focus on *qualitative* feedback. On the one hand, this has some disadvantages. For instance, it is not yet possible to generate qualitative real-time feedback merely automatically. However, it is not unlikely that this will be possible in the future. The quality of speech recognition is steadily improving and the automated analysis of arguments is also in constant progress (see e.g., Mu et al., 2012). On the other hand, we believe that mirroring qualitative information allows for richer feedback than solely mirroring quantitative information.

Type of Visualization The concept of group mirrors typically implies to provide information in a subtle and unobtrusive way. Our systems mostly comply with this. We only deviated from this in one project (see Chapter 8). Here, we provided textual feedback in one of the conditions, since our goal was to understand, how complex feedback speakers of a debate can process in real time.

Level of Aggregation The different levels of aggregation were investigated in Section 5.2. We compared three versions, one using solely *individual* feedback (i.e., using three balloons that represent the amount of ideas of three participants), one using solely *aggregated* feedback (i.e., one balloon representing the overall amount of ideas of the whole group) and a *mixed* version showing both information (i.e., showing the individual amount of ideas inside of a balloon representing the overall amount of ideas). That implies that the category "anonymity of the feedback receiver" was altered at the same time. Seeing individual feedback on a public display meant that information was *identifiable*, i.e., everyone could see the amount of ideas of every other group member. Showing only aggregated feedback meant that the displayed information was anonymous, as it was not possible to see, who contributed how many ideas. We summarized these characteristics as being of more *cooperative* nature (aggregated + anonymous) or more *competitive* nature (individual + identifiable).

Feedback Creation and Cardinality We used two main approaches to produce the feedback that then was visualized by the group mirror. For all of the group mirrors designed for creative tasks, we used the Wizard of Oz technique (see e.g., Kelley, 1983) to provide feedback about the quality of ideas. These tools are classified under the category *system*, as a Wizard of Oz experiment presents a tool in a simulation of what may be possible in the future even without human intervention.

The projects on argumentation and debates, in contrast, explicitly involve *humans* that provide the feedback. In both debate projects (see Chapter 8), a juror provides feedback to the speakers. In each moment, the juror provides feedback to one speaker exclusively. As in the course of a whole debate each speaker gets feedback from the juror, we still summarized these systems under the "cardinality" *one to many*. The prototype using the light cylinders (see Chapter 7) allows peers to give feedback to each other (i.e., the cardinality is *many to many*).

Anonymity of the Provider In two studies with the light cylinders (see Chapter 7), we compared an *anonymous* and an *identifiable* version. In the anonymous version, it was not possible for the group to detect who provided which feedback. In contrast, in the identifiable version, the feedback could be matched with the feedback provider.

Anonymity of the Receiver Not only the provider of the feedback but also the receiver can be *anonymous* or *identifiable*. Here, one can see that we also investigated the influence of more anonymous feedback, in contrast to related research that mostly focused on identifiable feedback. Investigating more anonymous forms of feedback arouse from the intention of designing group mirrors in a more pleasant way.

Placement and Privacy This intention of designing group mirrors in a more pleasant way that puts less pressure on the group members is one reason why we investigated also *private* and *semi-public* display environments. In the study described in Section 5.3, we compared a display environment using a private mode (i.e., showing individual information on private displays and aggregated information on a shared display) with a semi-public mode (i.e., showing individual information and aggregated information on a shared display). As already mentioned in the exemplary description of Figure 9.1, two public display settings (*table* and *wall*) were compared in the study with GROUPGARDEN (see Section 5.1).

Feedback Modality For this thesis, we exclusively investigated *visual* forms of group mirrors, for the reasons mentioned in the beginning of this section. To our best knowledge, all group mirrors from related work also use some kind of visual feedback.

Amount of Guidance Finally, most projects presented in this work can be classified as *mirroring* tools as they solely mirror certain information without showing how the ideal state would look like. In particular, systems that use identifiable information are all realized as mirroring tools. This decision was made since our goal was to reduce stress and pressure that might arise by showing identifiable information publicly together with a representation of the ideal state. Only systems that either display information in an aggregated way or show individual information on private devices use *metacognitive* feedback or *guidance*. In the study explained in Section 8.2, we compared a system showing information about the argument structure to a speaker of a debate. This can be classified as a metacognitive tool, as it shows the ideal state of an argument and how the speaker actually performed. We compared this to a version of that system that additionally shows textual information on how the speaker can improve. As this helps the speaker to reach the ideal state of an argument, this tool can be classified as a *guidance* system.

Two systems that did not compare different aspects of this design space were the two projects aiming at supporting the Disney Method (see Chapter 6). Both systems were exploratory and investigated the aspect of feedback that focuses more on the task at hand than explicitly on group processes. In the first study (see Section 6.1), feedback about the amount of ideas of certain roles of the Disney Method was visualized on a table. Implicitly, the overall amount of ideas of the group was displayed in an *aggregated* way. This setup was compared to a baseline without feedback in a study in the wild. The second study (see Section 6.2) was conducted with an interactive system showing feedback about the distributions of the Disney roles. Two versions were compared. One version served as a baseline and only showed the ideas that were stated by the group, colored according to the roles that were used. The second version additionally visualized which roles were underrepresented. As the ideal state is shown, this system is also classified as a *metacognitive* tool.

9.2 Discussion

Results of our studies revealed a number of benefits of group mirrors, some confirming results from related work, others providing novel insights into the influence of these systems. However, we observed that concept and design of these systems have a strong influence on the expectable effects. It seems inevitable to consider certain tradeoffs, for instance, between a possible performance increase and the acceptance of the group mirror. Moreover, certain application areas, such as learning, revealed to be more promising for the usage of group mirrors than others. I will, furthermore, briefly discuss our experiences with different evaluation methods for group mirrors.

9.2.1 Considering Tradeoffs

Based on the review of related work and our experiences with the different prototypes presented in this thesis, we conclude that designing group mirrors rather means to consider the use case and the goals of a group mirror carefully than to try to find a "perfect" group mirror that fits all situations and purposes. In this thesis, we particularly shed light on the display environment, two different use cases (creativity and argumentation), the concept underlying group mirror visualizations and the identifiability of the feedback provider. Results mainly revealed a tension between performance and acceptance, which I discuss below.

The Influence of the Display Environment

We compared two different display environments in two studies, both in the context of brainstorming.

Wall vs. Table A comparison of a group mirror on a table to one on a wall revealed that both display environments are suitable ways to improve brainstorming (see Section 5.1). While a group mirror on a wall was perceived as less disrupting and producing less pressure, the group mirror on the table gave a better sense of easy communication and collaboration and facilitated eye contact.

Implication: Depending on the usage scenario, both versions are reasonable possibilities. For instance, a group mirror on a wall is better suited for situations in which a group needs to focus on a complex task, while tasks in which communication plays a central role could be better supported through a group mirror on a table.

Public vs. Private The comparison of public and private display environments indicated that participants felt more comfortable, less under pressure and more motivated using private displays (see Section 5.3). Figure 9.2 depicts these results in a diagram. The axis labeled with *more private* represents the continuum of display environments from more public to more private. The two conditions of our study are indicated with *"public"* and *"private"*. The axes labeled with *more cooperative* and *more identifiable* will be discussed in the following sections. Conditions in which people felt most comfortable are connected with a yellow line (in this case the *"private"* display setup), conditions that produced the best performance are connected with a green line. The study did not reveal significant differences between the public and private conditions regarding performance. This is indicated with a *"?"* in the diagram and is the reason why the green line does not form a proper triangle.

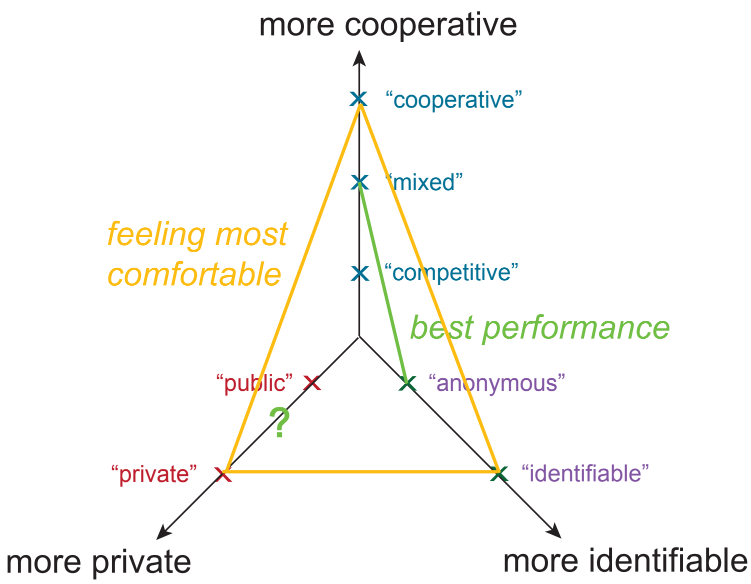


Figure 9.2: Overview of results. Results of three studies with three different prototypes regarding the aspects of how comfortable people feel (indicated in yellow) and in which conditions they achieve the best performance (green). Studies investigated the cooperativeness of the visualization (indicated in blue), the identifiability of the feedback provider (purple) and the privacy of the display environment (red).

Implication: As we could not find any performance-related differences, the deployment of private displays might be a reasonable alternative to the traditionally mostly public display environments used for group mirrors.

The Influence of More Cooperative Visualizations

We were interested in the question, whether the traditional form of group mirrors that mainly are designed in a more competitive manner (i.e., group members can compare their performance to the performance of the other group members) can be even improved in terms of performance and acceptance of the system by using more cooperative forms of visualizations (see Section 5.2). Therefore, we developed two novel concepts of group mirrors using cooperative and mixed visualizations.

We compared three versions, a competitive version, in which group members can compare their amount of ideas of a brainstorming session to the amount of ideas of the others, a cooperative version, in which only the overall amount of ideas is visible, and a mixed version that shows both, the individual amount and the overall amount of ideas. Results of this study revealed that performance was best in the mixed condition. Figure 9.2 depicts this by using a green line that connects the aspects in which we observed the best performance. One explanation for that could lie in the different personality traits that are best addressed in the mixed condition. While more extrovert and more agreeable people might benefit from the cooperative part, more introvert and less agreeable people might benefit from the competitive part of the mixed visualization. In terms of the well-being of participants, they reported to feel most comfortable in the cooperative condition (see Figure 9.2, yellow line), followed by the mixed condition, and the least comfortable in the competitive condition.

Implication: These results suggest that the effects of group mirrors in terms of performance can be improved by combining representations of individual performance with representations of the aggregated performance of the whole group.

The Influence of the Identifiability of the Feedback Provider

In two studies, we compared the influence of the anonymity or identifiability of the feedback provider in a collaborative argumentation scenario using peer feedback. In the first study, the identifiability was stronger, as the objects that displayed the feedback (light cylinders) were placed in front of the group members, while in the second study all cylinders were placed in the middle of the table with names written on the cylinders to identify the feedback provider (see Section 7.1). Results of the first study clearly showed a performance increase (i.e., more arguments were built) when feedback was provided anonymously in comparison to the condition in which the feedback provider was identifiable (see Figure 9.2). This might be attributed to the tendency that more feedback and also more negative feedback was provided in this condition. Moreover, the quality of the arguments was better in the anonymous condition compared to the identifiable condition. The second study partly confirms these results, since a tendency of more arguments and more arguments of higher quality could also be observed here (see Section 7.2). At the same time, however, participants did not feel as comfortable in the anonymous condition and also perceived the light objects as less valuable, especially in the first study.

Implication: These results suggest that technologically mediated peer feedback actually is an adequate possibility to improve argumentation and presumably also an opportunity to support people in learning how to argue. However, we could only show that this is the case when feedback was provided anonymously, albeit this was the condition in which group members felt the least comfortable.

A Tradeoff between Performance and Well-Being

One of the outcomes of the just described studies is that group mirrors can increase performance, however at the cost of the aspect, how comfortable people feel and how well the system is accepted by the group. The results regarding performance and well-being of participants are displayed in Figure 9.2. One needs to be aware that there might be interaction effects between these factors. In our studies, we only evaluated certain combinations of these aspects. Still, the information in this diagram might help to conceptualize group mirrors for different use cases. I will explain in the following, how the diagram depicted in Figure 9.2 may be used to accomplish this.

Designing group mirrors so that their triangle connecting the different aspects covers a larger area (e.g., a triangle connecting the points *"cooperative"*, *"private"* and *"identifiable"*) seem to be beneficial for the well-being of participants. Thus, this type of group mirrors might be especially suitable for inherently stressful situations. For example, in situations in which group members do not know each other, it might be more beneficial to introduce a group mirror that focuses on the overall performance of the group (i.e., using a more cooperative concept). This might support the group members in forming a new group and might assist them to reach a feeling of cohesiveness faster. Visualizing the feedback on private displays might take away another source of stress. In the case that feedback is provided by peers, it might be conducive to use identifiable feedback. This enables the feedback receiver to relate the feedback to a person who can add more nuanced information to the feedback through mimics and body language.

Group mirrors which are represented by a smaller triangle (e.g., connecting the points *"competitive"*, *"public"* and *"anonymous"*) might be particularly useful for situations that aim at increasing performance. Though we cannot make a statement about the performance in public and private display environments, the two other aspects (cooperativeness and identifiability) indicate that well-being and performance are divergent aspects, at least to some degree. Situations in which a performance increase is accompanied with less comfortable conditions is an issue inherent to a variety of scenarios. In sports for instance, high performance is achieved by considerable effort. Sportive competitions often create stressful situations that incite high performance. Education can be cited as another example. Here, concentrated learning can be strenuous, but is necessary for a deeper understanding of a topic.

For a large number of use cases, a tradeoff might be the best choice. It might, for example, be enough to choose one factor that is located in the outer area of the diagram. For example, using a cooperative visualization might be as agreeable for a group when it is shown on a public display as when it is depicted on private displays. As a cooperative visualization shows information in an aggregated way, the feedback cannot be attributed to the individual group members, which makes it less critical to show this information on a public display.

9.2.2 Benefits of Group Mirrors

Apart from the just described contemplations on the tradeoffs in terms of performance, acceptance and well-being, we can report about a number of clear advantages of group mirrors. A number of these have already been described in related research. By mirroring quantitative information, studies showed that group mirrors can increase speaking times (see e.g., Sturm et al., 2007) or balance participation (see e.g., Bergstrom and Karahalios, 2007b, 2009; DiMicco and Bender, 2007; DiMicco et al., 2007). By shifting our focus to qualitative information and other forms of feedback generation, such as peer feedback, our results add to the understanding of the effects of group mirrors on collaboration. Furthermore, we analyzed two specific use cases (collaborative creativity and argumentation) in detail. The main benefits of group mirrors that we observed in our studies are summarized below.

Supporting Groups with Qualitative Feedback

Group mirrors can make processes and information apparent that otherwise easily stay unnoticed. They may, for instance, reveal information about quantitative aspects of collaboration, such as speaking times or the amount of gazes in a certain direction (see Section 2.2.2). Our group mirrors, in contrast, focus on revealing qualitative aspects, for example, the amount of valid ideas (not solely speaking times) (see Chapter 5 and 6), information about the opinions or estimations of the other group members (see Chapter 7) or the assessment of a person outside of the group (see Chapter 8).

Our studies show that qualitative feedback can be valuable and may effectively support co-located collaboration. We can confirm results of related research (see e.g., Bergstrom and Karahalios, 2007c, 2009; Sturm et al., 2005) that showed that group mirrors may lead to an increased awareness of group processes. For example, group mirrors can make participants more aware of the performance of the other group members (see the projects presented in Chapter 5). Besides that, we show that feedback that is normally provided after a specific collaborative situation (such as an assessment of the quality of a debate) can effectively be included in real-time by mediating it through technology (see Chapter 8). Though participants partly reported that group mirrors may be distracting, they still managed to perform their main task while perceiving qualitative feedback and incorporating this feedback in real-time. Moreover, our results show that qualitative feedback mediated through group mirrors can influence the quality of the work (see e.g., Chapter 7 and 8).

Supporting Collaborative Creativity

Chapter 5 and 6 present group mirrors that are designed to support collaborative creativity. We chose two creativity techniques, first brainstorming, a well-known creativity technique, and second the Disney Method, which is a role-based creativity technique. For these tasks, results reveal that group mirrors can effectively help groups to remember and apply the specific rules of creativity techniques. The brainstorming rule "no criticism", for instance, was successfully established when a group mirror was present (see Chapter 5). In addition, participation tended to be more balanced with the assistance of group mirrors, an aspect that has proved to be beneficial for brainstorming sessions (Oxley et al., 1996).

The concept of the Disney Method to use the three roles *dreamer*, *realist* and *critic* was also supported by the feedback of group mirrors (see Chapter 6). Results indicate that especially unpopular roles such as the *critic* were used more often when a supporting system was used. A group mirror can, moreover, increase motivation and raise the amount of ideas that a group produces (see Section 6.2). Besides that, when we asked participants explicitly which conditions they liked better, a majority of participants preferred the sessions with support of a group mirror compared to the sessions without such a support.

Supporting Argumentation

In addition, we designed group mirrors aiming at supporting argumentation. Again, two tasks were analyzed in more detail. One of these tasks was collaborative argumentation (see Chapter 7), the other one was argumentative debate (see Chapter 8). For collaborative argumentation, results indicate that, when the group mirror is designed in an adequate way (i.e., providing feedback anonymously), groups build more arguments and also arguments of higher quality (see Section 7.1). Furthermore, participants reported about a number of positive experiences with the light cylinders that allowed them to provide peer feedback to the group. For instance, they liked that they were able to get an overview of the opinions of the other group members quickly, they valued that everyone can contribute at the same time and perceived the system as stimulating. The prototypes designed for argumentative debates were liked especially by novices and experts (see Chapter 8). When the feedback of the system is designed adequately (i.e., using feedback about the argument structure), results reveal the tendency that speakers build more arguments and also arguments of higher quality.

Supporting Groups with Content-Related Feedback

We conducted a number of studies in which the feedback did not only mirror collaborative processes but also included information about the content of the group work. Two of these projects have the goal to support the Disney Method. The first one mirrors information about the role distribution (i.e., the distribution of the roles of the Disney Method: *dreamer*, *realist*, *critic*) to the group using a subtle visualization on a tabletop display. This visualization mirrors information about the activities of the group (number of ideas of the whole group) next to information about the content (role distribution). Results show that the overall amount of ideas increased when using this group mirror compared to a baseline.

For the second project, we developed an interactive system running on tablets in which group members could enter their ideas and choose the different roles of the Disney Method for that purpose. In one version, we added a functionality to show the balance of the roles. In this study, we could see a tendency that not only the different roles were used in a more balanced way but also that *below average* participants increased their participation with the system.

These results indicate that mirroring aggregated information about the group performance (e.g., number of ideas of the whole group) together with more content-related aspects (e.g., role distribution) can have a similarly positive effect on group processes as more traditional group mirrors that show individual information (e.g., number of ideas of each group member) without additional content-related information.

9.2.3 The Different Types of Feedback Creation

For the different prototypes, we used three main methods to create the qualitative feedback. Our experiences with these are discussed below.

Feedback from a Wizard of Oz

For creative tasks (see Chapter 5 and 6) we used Wizard of Oz experiments (see e.g., Kelley, 1983). The experimenter took over the task of the "wizard" and provided feedback, for instance, about the amount of ideas. This worked well for our studies. However, it is not yet predictable by when this kind of real-time qualitative analyses will be achievable automatically. If there is no trained person present, who can take over the role of the "wizard", an alternative is to make use of peer feedback.

Feedback from Peers

Our experiences with peer feedback reveal that it might lead to reservations against the feedback (see Chapter 7). Research on the effects of computer-mediated and person-mediated feedback claim that people do not perceive feedback from a system as evaluative and aversive as person-mediated feedback (see also related work about this topic in Section 2.2). It might be possible that the same effects as with person-mediated feedback occur with feedback from a person, mediated through technology, as it is the case in our prototypes. This could explain why the group mirrors using seemingly computer-mediated feedback (by using the Wizard of Oz technique), were liked better than the group mirrors providing feedback from a person, mediated through technology. Still, results regarding the effects of peer feedback in terms of performance and quality of the group work reveal the potential of this method.

Feedback from an Independent Person

The third option that we evaluated was feedback provided through a person outside of the group. In our case, a juror provided feedback about debating performance to the speakers of a debate (see Chapter 8). The feedback from this person was accepted well. We assume that this can be attributed to the situation that people are used to get feedback from a jury. In contrast to the feedback through peers, participants were aware of the special role of the juror, who is a person with considerable more experience in debating than most of the speakers. Of course, a teacher or moderator can also take the role of the independent person, reliant on the usage scenario.

We conclude that at this point, feedback from an independent person might be a profitable choice in many situations. Feedback from such a person is of high quality and groups might rather accept it than feedback from peers. Still, if there is no adequate person available, feedback from peers can be a valuable alternative, as it can achieve similarly good effects on the quality of the work.

9.2.4 Application Areas

As outlined before, group mirrors can have a number of benefits for collaboration. However, introducing group mirrors in collaborative activities also comes with a burden. In all studies, participants mentioned that the feedback sometimes distracted from the main task. Moreover, participants referred to feel pressured by the feedback. Especially when feedback was provided by peers, problems such as artificiality of the situation, the possibility of bullying, streamlining of the discussion, the loss of subtle cues and the granularity of person-mediated feedback were mentioned. This raises the question, in which situations these negative effects of group mirrors might be tolerable and in which not.

Based on the studies that we conducted, we estimate that subtle and unobtrusive group mirrors might be suitable for constant support of collaboration. The group mirror presented in Section 6.1 uses a subtle visualization composed of simple shapes and colors shown on a peripheral display, mirroring aggregated information. This is an example of a group mirror that blends in the environment and does not interrupt group work much. The other extreme is a group mirror using light cylinders as we presented in Chapter 7. Group members actively operate the system. This seems less suited for using it as a constant support during discussions, as it might impede a fluent conversation. This system might be more appropriate for an educational setting. Through encouraging group members to generate feedback by themselves, learning about the tasks they perform may also be facilitated.

In summary, we believe that the context of use is rather *learning how to conduct a certain collaborative technique* (e.g., *learning to argue*, *learning to brainstorm* etc.) than using these systems as a constant support during collaborative work, as we could observe an increase in performance and quality of the work by group mirrors, but additionally observed the two main problems of disruption and pressure.

9.2.5 Evaluating Group Mirrors

We investigated a number of different methods to design and evaluate our studies. Our experiences with laboratory studies and field studies and different evaluation methods are outlined in the following.

Laboratory Studies

Generally, we designed the comparative studies in a very controlled way and tried to eliminate as many confounding variables as possible. On the one hand, this allowed us to identify some effects that specific aspects of the design space have on group processes. On the other hand, this sometimes resulted in artificial situations. Moreover, altering only one variable of the characteristics that we identified in the design space at a time can mean such a small difference that strong results are not to be expected. This problem can be described for the example of the light cylinders. Here, we conducted a first study in which the placement of the light cylinders can be seen as a confounding variable. This, however, strengthened the identifiability of the feedback. We were able to reveal a number of significant differences between the conditions. In the second study, we eliminated this confounding variable, leading to a situation in which the identifiability was less prominent. Results only revealed tendencies, statistically significant differences were not observed. This problem has also been described in a similar context by Streng (2012).

Field Studies

We performed two field studies. The first field study was conducted during a practical course at a university (see Section 6.1). Group members had to discuss ideas of their colleagues using the Disney Method. The group mirror showed the anticipated effects, more ideas were generated and the roles were used in a more balanced way. The majority of the participants stated that they liked the sessions with group mirror more than those without.

The second study was conducted in the course of the regular debates of a debate club (see Section 8.1). Here, we noticed some initial skepticism against using technologically mediated feedback in the formerly analog process, although, afterwards, most of the group members felt that the system helped to mediate the feedback of the juror so that the speakers could incorporate it in real-time. Novices and experts found the feedback helpful, but only when they considered the general structure that the system suggested as helpful.

These two studies offer a first glimpse on how group mirrors might be adopted in real collaborative situations. To get an even better understanding on how group mirrors are used in real situations, these systems would additionally need to be given to groups to use them over a longer period of time. That is also my main suggestion for future work on group mirrors. Other suggestions and limitations of our current approach will be discussed in Section 9.4.

Evaluation Methods

As the projects are not necessarily reported in chronological order, it is difficult to detect how our methods have changed over time. For evaluating the balance of participation, for example, we started by using a categorization of *below average* and *above average* participants and analyzed, how the amount of contributions of these two groups changes. This is also a method that has been used in related research (see e.g. DiMicco et al. (2007)). In one of the following studies, we gave standard deviations to evaluate more or less balanced sessions a try (i.e., high standard deviation in the amount of contribution indicates less balanced participation). Finally, we switched to calculating the Gini coefficient (Weisband et al., 1995), a measurement that has also been used in related research on group mirrors (see e.g. Schiavo et al. (2016)). To conclude, we consider all of these different methods as valid and useful tools to estimate the balance of participation. However, for reasons of comparability, we would suggest to rather stick to already established methods, such as using the categorization of *below average* and *above average* participants or the Gini coefficient, as we could not detect further advantages of the method using the standard deviation.

We also investigated the suitability of physiological measurements for evaluating group mirrors. In one study (see Section 8.2), we measured heart rate and affect in speech. The results from the heart rate did not reveal additional insights. Speakers had a higher heart rate during their speech than before and after, however, independent from the presence of a group mirror. It might be possible that BCIs (Brain Computer Interfaces) provide further information. The measurements of the affect in speech did reveal some interesting insights, thus, it might be interesting to deploy it in future studies on group mirrors.

For analyzing quantitative results, we used tests that take into account that participants were nested in groups, when necessary. We mostly used established tests such as a nested ANOVA or linear mixed models. However, as concerns with null hypothesis significance testing are growing in various fields (Cumming, 2013; Dragicevic, 2015), we started to report effect sizes with 95% confidence intervals (CI). This was done in the study on debates, in which participants were not nested in groups. In our opinion, reporting confidence has a number of advantages. For instance, confidence intervals can present the same information as *p*-values and are at the same time easier to understand (Dragicevic, 2015). However, we experienced reservations towards this method in the field of HCI.

9.3 Limitations

The main limitations of this work can be found in the study designs and consequently the generalizability of the results. I will also discuss possible risks that the employment of group mirrors can bring along. Finally, the term *group mirror* that we used throughout this thesis is reviewed briefly.

9.3.1 Study Designs

Limitations regarding the study designs include size and structure of the study samples, the target groups, the well-known issues of laboratory studies and the limitations of short term evaluations.

Size and Structure of Study Samples

One main limitation of the laboratory studies presented in this thesis is the small amount of groups that took part in the experiments. As the topic of this work was to evaluate collaborative activities, all studies were conducted with groups of three or four participants. However, finding and coordinating participants in groups of up to four people is often challenging. Several studies were conducted in the course of Bachelor or Master theses, leading to additional time constraints. Consequently, some of the laboratory studies were conducted with a relatively small number of participants that were nested in groups of three or four participants (reaching from 18 participants nested in 6 groups to 48 participants nested in 12 groups). Moreover, group dynamics can arise within these groups that can influence the results. The familiarity of group members or the different personalities of participants might vary. It is also likely that the amount of dominant and non-dominant persons in each group differed.

We counteracted some of these problems. For instance, all group members within each group already knew each other before the studies. That ensured that the phase in which people get to know each other was skipped in all groups. The familiarity might still vary, dependent on how long group members know each other or if they are related to each other. In addition, we evaluated the studies using adequate statistical methods based on the participant level (e.g., nested ANOVA or a linear mixed model with groups as a random intercept that take into account that participants are nested within groups). Still, results of the studies often revealed only tendencies.

The difficulty to conduct studies with a large amount of participants also was a reason to decide in favor of choosing relatively small groups. In all our laboratory studies, groups were formed of three or four participants. Though our group mirrors are intentionally designed for smaller groups, it is still difficult to estimate, how groups with more than just three or four participants respond to group mirrors.

Target Groups

Another issue is the target group of our laboratory studies. While in our largest study (the first study with the light cylinders, see Section 7.1), we managed to include participants with a variety of different professions of all ages (19 to 76 years) and with the same amount of female and male participants, this was not achievable for the other studies. For most studies, participants were recruited at our university, and a majority of them were media informatics students. Thus, a large number of our participants came from Germany, were young, technology-savvy and highly educated. This clearly does not reflect an average of the population. Still, it might fit our target group for group mirrors, as they are designed to support people who are interested in learning how to use creativity techniques or to learn how to lead a discussion in a structured way.

Issues of Laboratory Studies

As we conducted mainly laboratory studies (except of two studies in the wild), the well-known characteristics and problems of these kinds of studies also apply for our studies. These are, amongst others, an increased internal validity at the expense of less external validity, artificiality of the situation or the effect of demand characteristics, meaning that participants behave, according to how they think they are required to behave. For instance, in the projects on brainstorming (see Chapter 5), rules such as "do not interrupt each other" or "stay on topic" had to be observed. Because of the artificial situation in the laboratory study and the awareness of being observed, group members probably took more care not to break these rules than in a regular brainstorming session.

Missing Insights into Long-Term Effects

Finally, all our studies have been conducted over a short period of time. Participants mostly took part in the studies for about an hour. Only the study with the debating club was carried out on two different dates with one week in between. This made it impossible to identify long term effects. In this regard, it would be interesting to know, how systems are used after the novelty effect has faded away, if there will be long term learning effects or if users internalize the effects of the group mirrors after a while.

9.3.2 Generalizability of Results

Due to several of the just described limitations, the generalizability of the results is constrained. Studies were conducted with mainly young people, with only small groups and with group members that are familiar with each other. In addition, we specifically chose two tasks: collaborative creativity and argumentation. We still regard our considerations for the design of group mirrors that we outlined in the different chapters and in Section 9.2 as a valuable help for designing group mirrors and as a basis for future research. However, further studies will be necessary in other usage contexts to gather more insights.

9.3.3 Possible Risks of Group Mirrors

Related research and the various studies presented in this thesis proved that group mirrors have an influence on how people behave in collaborative situations. They can make group members aware of processes that are otherwise difficult to perceive and thus, achieve that people reflect on their behavior and their position within the group. However, this does not mean that the effects of group mirrors are positive by default. For instance, Bergstrom and

Karahalios (2012) were able to show that participants also react to skewed visualizations and adapt their behavior accordingly. On the one hand, this brings the risk that people misuse group mirrors to change a group's behavior to their demands. On the other hand, unwanted effects of group mirrors can happen unnoticed and with no ill intent. Thus, designers of group mirrors need to consider their choices very carefully in terms of how to design group mirrors, which effects are possible and which are desired or not.

9.3.4 Terminology

We used the term *group mirror* throughout the thesis as an umbrella term for all systems that we presented. Thereby, we stretched the original definition by Jermann et al. (2001). We included the possibility that feedback can be provided by peers or other persons. We also discussed different cardinalities: *one to one*, *one to many*, *many to one* and *many to many*. In the study on debates, for instance, one juror provides feedback to the speakers of a debate. As the speakers have clearly determined speaking slots, the juror only provides feedback to one speaker at a time. We still classified this system as a group mirror, as debating is a collaborative activity and speakers receive feedback about their collaborative activities (e.g., in the version with textual feedback, how well they answered to an argument stated before). Still, an also appropriate terminology for describing these kinds of systems is *technologically mediated feedback*, which better captures the possibility that also persons may provide the technological feedback. However, we decided to use the established term *group mirror* in this thesis, as we strongly built on previous research on group mirrors.

9.4 Future Work

There are various ways to proceed with the work on group mirrors starting from related research and the present thesis. In the following, I will discuss other possible use cases for group mirrors, factors of the design space that have been widely neglected until now and feasible follow-up studies that succeed from the studies presented in this thesis. More importantly, I will discuss the directions for future research that seem most promising in my eyes.

9.4.1 Use Cases

Group mirrors might be employed in a variety of different scenarios. We suggest three areas, which we assume as promising directions for future investigations.

Education

After having conducted a number of studies, we estimate the main usage scenario of group mirrors in *learning how to conduct a specific collaboration technique.* As already discussed in the previous chapter, we observed that group mirrors can help to increase group members' performance and also the quality of their work immediately. However, incorporating group mirrors can also disrupt and distract from the main task. Thus, we estimate that one main use case for group mirrors bears on education, be it in school, university or adult education, with the goal to learn how to apply specific collaborative techniques (e.g., *learn to argue*, *learn to brainstorm*).

Meetings

We also thought about different other use cases apart from brainstorming and argumentation. One idea is to employ group mirrors in business meetings. People complain about meetings that take too long, are structured badly or are dominated by certain persons. Especially when meetings are not moderated, group mirrors might help to give meetings a structure. In that context, a number of other factors probably are of importance, for instance, hierarchies between the group members that are often preexisting in a business context.

Qualitative Analyses

For supporting researchers, we thought about integrating awareness information in qualitative analyses, a task that is often conducted collaboratively. Here, people frequently switch between co-located collaboration and remote collaboration, as the involved persons are in many cases situated in different locations. Furthermore, current methods of qualitative analyses often include several switches between the physical space (e.g., paper-based methods) and the digital space, thus bringing along a number of new challenges. We are currently working on a research project on this topic together with the InnoVis Group of the University of Calgary.

9.4.2 Unregarded Factors of the Design Space

The design space that was presented in this thesis shows some areas that have been mostly neglected by research on group mirrors. In my opinion, three areas are especially interesting for future investigations.

Group Composition

One of these suggestions is to investigate the *group composition* in more detail. Factors, such as the type of the group, the group size or the familiarity of the group members probably have an influence on how the system should be designed and how it is used and accepted. For instance, a group of students that employs a group mirror to learn how to use a creativity technique might be more content with a playful system. Our project on GROUPGARDEN (see Section 5.1) implements such an approach and represents group members using flowers and trees. Feedback from participants confirmed that the group mirror seemed better suited for a group of students than for a group of persons in a professional setting. Playful approaches might also help groups that are newly formed to ease the situation. In professional contexts, such as business meetings, group mirrors might need to take into account existing hierarchies. Moreover, it might be interesting to investigate, how group mirrors should be designed for larger groups, for example, by finding suited ways to aggregate information. When a large group is represented on a display, it might be impossible to relate the information to a specific person. Thus, aggregating the information in a proper way might be a solution to this issue.

Mixed Presence Scenarios

Moreover, I estimate situations of *mixed presence* (a mix of co-located and remote participants) as an interesting use case for group mirrors. For group members that are cut in remotely it is, for example, difficult to take the floor, as their ability to indicate that they want to talk (e.g., through body language) is limited. An idea is to introduce group mirrors that represent remote group members to the co-located group in a way so that they can be integrated more effectively. The light cylinders presented in Chapter 7 could be used for that purpose, as their information can easily be perceived in the periphery of the attention by the co-located group members. Thus, a remote person can call the attention to him- or herself, without interrupting the conversation verbally.

Auditory and Tactile Feedback

Finally, *auditory* and *tactile* feedback might be suitable possibilities to include in group mirrors. Auditory feedback could be a valuable alternative to visual feedback if the main task includes to conduct work in which the group visually focuses on other material. In the second phase of brainstorming, the *norming* phase, people often use pen-and-paper methods to create mind-maps to cluster their ideas. Integrating a visual feedback tool into this process might be challenging. An alternative could be auditory feedback, for instance, in form of background sounds that change smoothly dependent on the ongoing group processes. Tactile feedback has the additional advantage that other group members stay unaware of the feedback that one perceives. A vibrating bracelet might be used to make group members aware of group processes. Group members might, for example, be alerted when it is reasonable to yield precedence to others.

9.4.3 Possible Follow-Up Studies

Based on the results of the studies presented in this thesis, I can suggest a number of follow-up studies.

Further Combinations of the Characteristics of Group Mirrors

The two studies adapting the balloon representation that compare public and private displays and more cooperative, mixed and competitive visualizations, suggest that performance increases when applying mixed visualizations while group members feel most comfortable using the cooperative version, but also using the private display environment (see Chapter 5). These results indicate that the combination of private displays and cooperative visualizations could increase the well-being of participants and a combination of private displays and mixed visualizations could result in better performance. However, these assumptions are not yet validated and another study investigating these combinations could reveal further insights into this issue.

Considering Personality Traits

One attempt to explain the results of the study on cooperative and competitive visualizations was that people with different personality traits react differently to group mirrors (see Section 5.2). Evaluating this relation further could reveal interesting results. This might be accomplished by retrieving information about the personality of participants before the study by using the big-five personality dimensions (see e.g., Digman and Inouye, 1986; Fiske, 1949), for example. This could, furthermore, create the opportunity to design adaptable group mirrors that can be adjusted dependent on the personalities of individual group members.

Comparison with a Moderator

Most studies presented in this thesis compare one or more conditions to a baseline condition without technologically mediated feedback. Results show that group mirrors have a number of advantages compared to these baseline conditions. However, it would be interesting to evaluate, if group mirrors also have advantages compared to group work with a trained moderator. Group mirrors can provide feedback in a very unobtrusive and subtle way. Then again, a trained moderator can probably adapt to new and unknown situations in a much more flexible way. However, differences, advantages and disadvantages could best be evaluated in a study.

Of course, a large amount of studies like this (e.g., comparing content-related with participant-related feedback, comparing quantitative and qualitative feedback or comparing face-to-face feedback directly from peers with technologically mediated peer feedback) are conceivable and would probably lead to further insights. However, I think that the next steps of research on group mirrors are not to investigate every other aspect of the design space. I rather think that next steps should be to conduct long-term studies and to analyze if it is possible to establish group mirrors in “real” contexts. I will discuss these issues in the next section.

9.4.4 Directions for Future Research

Besides the important next step of conducting long-term studies in real usage scenarios, I discuss the aspects of collaborative performance tracking and general ideas on how existing technologies that are present in collaborative situations could be used as group mirrors.

Long-Term Studies

Multiple studies from related research together with the work presented in this thesis have led to a better understanding of the effects that group mirrors have on group processes. However, what is still missing is a comprehension of the effect of group mirrors on collaboration over a longer period of time. It would be interesting to know, if group mirrors actually can serve as constant support during collaboration, or if groups will internalize the feedback of the group mirror so that it will be superfluous after a while.

Real Use Cases

Insights are missing on how group mirrors will actually be applied in real usage scenarios. To our best knowledge, group mirrors are widely unknown outside of the research community. This could have several reasons. Maybe, people just do not know of such systems. This is quite likely, as to our best knowledge, there are only few commercial group mirror systems (e.g., classroom response systems). Moreover, there might be an initial inhibition level that needs to be overcome. Thus, an important next step is to try to establish group mirrors in real contexts and observe their effects over a longer period of time.

Group Mirrors as Moderator Support

Apart from these two suggestions, another direction for future research in this area might be worth to investigate: building systems that do not support the group directly but are designed to assist a moderator. This would have the advantage that a moderator could be made aware of group dynamics that are difficult to perceive without the support of a group mirror. Then, the moderator could decide, in which moments it makes sense to use this information to support the group. The group might feel less disrupted and less under pressure, as information about their behavior are not always visible. In addition, such a tool could also be used to train moderators.

Tracking Collaborative Data

Tracking personal data is a trend that is recently growing in popularity (see e.g., Epstein et al., 2014; Rooksby et al., 2014). For example, people track their sportive activities, their sleep or nutrition information. Apps and services are used to track individual performance. The collaborative part is mostly achieved through communities, in which people can compare their information to the information of other persons. While tracking performance of groups is in no way novel (e.g., in group sports), using systems such as group mirrors to track and visualize group performance is not yet a widespread method. Comparable existing approaches are competitive activities between different teams of companies. These compare their sportive activities, for instance, tracked by step counters. Tracking collaborative activities instead or in addition to individual personal data can enable new opportunities in these cases of application.

Making Use of Existing Technologies

Introducing group mirror systems into co-located collaboration can cause that a former analogous situation now is supported with technology. Then again, this change has already taken place in a majority of collaborative situations. People bring their laptops, smartphones, smartwatches or, lately also head-mounted displays, which changes the culture of conversation and collaboration. It can lead to situations in which people are less involved in the group work as they are engaged with other activities on their devices, be it work-related or private. Group mirrors might constitute a way to use this existing technology to enhance collaborative processes. “Converting” personal devices into group mirrors for the time of the group work might help to reduce distracting activities with the additional advantages that group mirrors bring with them.

Collaboration is such an important aspect of our daily lives that research constantly needs to try to understand group processes better and, as a second step, to develop and enhance methods that aim at supporting collaboration. Group mirrors are one approach that tries to achieve this complex endeavor. The findings from related research and the results presented in this thesis provide considerable evidence that the support through such subtle systems can actually change group work to the better. In my eyes, the next step must now be to employ group mirrors in real usage scenarios over a longer period of time. I hope that with this thesis, I could shed light on the different characteristics, effects and possible use cases of group mirrors and pave the way for directions of future work in this area.

BIBLIOGRAPHY

H. Adachi, S. Myojin, and N. Shimada. 2014. Tablet system for sensing and visualizing statistical profiles of multi-party conversation. In *IEEE 3rd Global Conference on Consumer Electronics (GCCE)*. IEEE, 407–411.

H. Adachi, S. Myojin, and N. Shimada. 2015. ScoringTalk: A tablet system scoring and visualizing conversation for balancing of participation. In *SIGGRAPH Asia 2015 Mobile Graphics and Interactive Applications (SA '15)*. ACM, Article 9, 5 pages.

J. A. Adams. 1978. Theoretical issues for knowledge of results. In *Information processing in motor control and learning*. Motor Behavior Laboratory, University of Wisconsin-Madison, Academic Press, 229–240.

M. J. Adams, Y. J. Tenney, and R. W. Pew. 1995. Situation awareness and the cognitive management of complex systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1, 85–104.

M. Aiken, M. Vanjani, and J. Paolillo. 1996. A comparison of two electronic idea generation techniques. *Information and Management* 30, 2, 91–99.

S. Ainsworth, G. Gelmini-Hornsby, K. Threapleton, C. Crook, C. O'Malley, and M. Buda. 2011. Anonymity in classroom voting and debating. *Learning and Instruction* 21, 3, 365–378.

H. Alavi and P. Dillenbourg. 2012. An ambient awareness tool for supporting supervised collaborative problem solving. *IEEE Transactions on Learning Technologies* 5, 3, 264–274.

G. S. Alder and M. L. Ambrose. 2005. An examination of the effect of computerized performance monitoring feedback on monitoring fairness, performance, and satisfaction. *Organizational Behavior and Human Decision Processes* 97, 2, 161 – 177.

V. Aleven and K. D. Ashley. 1997. Teaching case-based argumentation through a model and examples empirical evaluation of an intelligent learning environment. In *Proceedings of the 8th International Conference on Artificial Intelligence in Education (AI-ED 1997)*. IOS Press, 87–94.

T. M. Amabile. 1982. Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology* 43, 5, 997 – 1013.

T. M. Amabile. 1983. The social psychology of creativity: A componential conceptualization. *Journal of personality and social psychology* 45, 2, 357–376.

T. M. Amabile. 1993. Motivational synergy: Toward new conceptualizations of intrinsic and extrinsic motivation in the workplace. *Human Resource Management Review* 3, 3, 185 – 201.

T. M. Amabile. 1996. *Creativity in context: Update to the social psychology of creativity."*. Westview press.

R. J. Anderson, R. Anderson, T. VanDeGrift, S. Wolfman, and K. Yasuhara. 2003. Promoting interaction in large classes with computer-mediated feedback. In *Designing for change in networked learning environments*. Springer, 119–123.

J. Andriessen. 2006. Arguing to Learn. In *Handbook of the Learning Sciences*. Cambridge University Press, 443–459.

J. Annett. 1969. *Feedback and Human Behaviour*. Penguin Books.

S. J. Ashford and A. S. Tsui. 1991. Self-regulation for managerial effectiveness: The role of active feedback seeking. *Academy of Management journal* 34, 2, 251–280.

C. S. Asterhan and T. Eisenmann. 2011. Introducing synchronous e-discussion tools in co-located classrooms: A study on the experiences of ‘active’ and ‘silent’ secondary school students. *Computers in Human Behavior* 27, 6, 2169 – 2177.

A. Attwenger. 2015. *Evaluating the impact of anonymous and identifiable feedback on argumentation quality*. Bachelor thesis.

K. Bachour, F. Kaplan, and P. Dillenbourg. 2010. An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Transactions on Learning Technologies* 3, 3, 203–213.

K. Bachour, F. Kaplan, and D. Pierre. 2008. Reflect: An interactive table for regulating face-to-face collaborative learning. In *Proceedings of the Third European Conference on Technology Enhanced Learning (EC-TEL 2008)*. Springer, 39–48.

M. Baker. 1999. Argument and constructive integration. In *Foundations of argumentative text processing*. Amsterdam University Press, 179–201.

M. Baker. 2003. Computer-mediated argumentative interactions for the co-elaboration of scientific notions. In *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments*. Springer, 47–78.

S. Bakker, E. van den Hoven, and B. Eggen. 2012. FireFlies: Supporting primary school teachers through open-ended interaction design. In *Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI '12)*. ACM, 26–29.

S. Bakker, E. van den Hoven, and B. Eggen. 2013. Fireflies: Physical peripheral interaction design for the everyday routine of primary school teachers. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction (TEI '13)*. ACM, 57–64.

A. Bandura. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review* 84, 2, 191–215.

C. I. Barnard. 1938. *The functions of the executive*. Harvard University Press.

F. Barron and D. M. Harrington. 1981. Creativity, intelligence, and personality. *Annual review of psychology* 32, 1, 439–476.

E. M. Barth and E. C. Krabbe. 1982. *From axiom to dialogue: A philosophical study of logics and argumentation*. Walter de Gruyter.

M. Batey and A. Furnham. 2006. Creativity, intelligence, and personality: A critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs* 132, 4, 355–429.

D. Baumgart, A. Pohl, V. Gehlen–Baum, and F. Bry. 2011. Providing guidance on backstage, a novel digital backchannel for large class teaching. *Education in a Technological World: Communicating Current and Emerging Research and Technological Efforts*, 364–371.

I. D. Beatty. 2004. Transforming student learning with classroom communication systems. *Educause Center for Applied Research (ECAR) Research Bulletin* 2004, 3.

B. Beersma, J. R. Hollenbeck, S. E. Humphrey, H. Moon, D. E. Conlon, and D. R. Ilgen. 2003. Cooperation, competition, and team performance: Toward a contingency approach. *Academy of Management Journal* 46, 5, 572–590.

A. D. Bergstrom. 2006. *Visualization of audio augmenting social interactions*. Ph.D. Dissertation. University of Illinois at Urbana-Champaign.

T. Bergstrom and K. Karahalios. 2007a. Conversation Clock: Visualizing audio patterns in co-located groups. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS '07)*. IEEE, 78–87.

T. Bergstrom and K. Karahalios. 2007b. Conversation Votes: Enabling anonymous cues. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems (CHI EA '07)*. ACM, 2279–2284.

T. Bergstrom and K. Karahalios. 2007c. Seeing more: Visualizing audio cues. In *Proceedings 11th IFIP TC 13 International Conference (INTERACT '07)*. Springer, 29–42.

T. Bergstrom and K. Karahalios. 2009. Vote and be heard: Adding back-channel signals to social mirrors. In *Proceedings of the 12th IFIP TC 13 International Conference (INTERACT '09)*. Springer, 546–559.

T. Bergstrom and K. Karahalios. 2012. Distorting social feedback in visualizations of conversation. In *Proceedings of the 45th Hawaii International Conference on System Science (HICSS '12)*. IEEE, 533–542.

M. Billig. 1996. *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge University Press.

J. H. Block. 1971. Criterion-referenced measurements: Potential. *The School Review* 79, 2, 289–298.

M. Brandon, S. Epskamp, T. de Groot, T. Franssen, B. van Gennep, and T. Visser. 2011. The effects visual feedback on social behavior during decision making meetings. In *Human Interface and the Management of Information. Interacting with Information*. Lecture Notes in Computer Science, Vol. 6772. Springer, 219–228.

J. D. Bransford, A. L. Brown, and R. R. Cocking. 1999. *How people learn: Brain, mind, experience, and school.* National Academy Press.

J. Brockner. 1979. The effects of self-esteem, success-failure, and self-consciousness on task performance. *Journal of Personality and Social Psychology* 37, 10, 1732–1741.

S. M. Brookhart. 2008. *How to give effective feedback to your students*. Association for Supervision and Curriculum Development (ASCD).

F. Bry, V. Gehlen–Baum, and A. Pohl. 2011. Promoting awareness and participation in large class lectures: The digital backchannel backstage. In *Proceedings of th IADIS International Conference e-society*. 27–34.

J. Buder and D. Bodemer. 2007. Supporting controversial CSCL discussions with augmented group awareness tools. In *Proceedings of the 8th Iternational Conference on Computer Supported Collaborative Learning (CSCL'07)*. International Society of the Learning Sciences, 93–101.

G. Burns. 1995. The secrets of team facilitation. *Training and Development* 49, 46–46.

J. E. Caldwell. 2007. Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education* 6, 1, 9–20.

J. Carifio and R. J. Perla. 2007. Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Sciences* 3, 3, 106–116.

P. H. Carstensen and K. Schmidt. 1999. Computer Supported Cooperative Work: New Challenges to Systems Design. In *Handbook of Human Factors and Ergonomics*. 619–636.

C. S. Carver and M. F. Scheier. 1981. *Attention and self-regulation: A control-theory approach to human behavior*. Springer.

C. S. Carver and M. F. Scheier. 1998. *On the Self-Regulation of Behavior*. Cambridge University Press.

W. Cassidy, C. Faucher, and M. Jackson. 2013. Cyberbullying among youth: A comprehensive review of current international research and its implications and application to policy and practice. *School Psychology International* 30, 4, 1 – 38.

C. P. Cerasoli, J. M. Nicklin, and M. T. Ford. 2014. Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin* 140, 4, 980–1008.

M. Chau. 2011. Visualizing web search results using glyphs: design and evaluation of a flower metaphor. *ACM Transactions on Management Information Systems (TMIS)* 2, 1, 27 pages.

M. Chen. 2003. Visualizing the pulse of a classroom. In *Proceedings of the Eleventh ACM International Conference on Multimedia (MULTIMEDIA '03)*. ACM, 555–561.

K. Cho and C. Schunn. 2007. Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers & Education* 48, 3, 409–426.

K. M. Christopherson. 2007. The positive and negative implications of anonymity in internet social interactions: "on the internet, nobody knows you're a dog". *Computers in Human Behavior* 23, 6, 3038–3056.

R. B. Clariana, D. Wagner, and L. C. Roher Murphy. 2000. Applying a connectionist description of feedback timing. *Educational Technology Research and Development* 48, 3, 5–22.

A. Clayphan, A. Collins, C. Ackad, B. Kummerfeld, and J. Kay. 2011. Firestorm: A brainstorming application for collaborative group work at tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '11)*. ACM, 162–171.

A. Clayphan, J. Kay, and A. Weinberger. 2014. ScriptStorm: Scripting to enhance tabletop brainstorming. *Personal and Ubiquitous Computing* 18, 6, 1433–1453.

E. G. Cohen. 1994. Restructuring the Classroom: Conditions for Productive Small Groups. *Review of Educational Research* 64, 1, 1–35.

T. Connolly, L. M. Jessup, and J. S. Valacich. 1990. Effects of anonymity and evaluative tone on idea generation in computer-mediated groups. *Management Science* 36, 6, 689–703.

T. Connolly, R. L. Routhieaux, and S. K. Schneider. 1993. On the effectiveness of group brainstorming test of one underlying cognitive mechanism. *Small Group Research* 24, 4, 490–503.

S. Consolvo, P. Klasnja, D. W. McDonald, D. Avrahami, J. Froehlich, L. LeGrand, R. Libby, K. Mosher, and J. A. Landay. 2008a. Flowers or a robot army?: Encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08)*. ACM, 54–63.

S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay. 2008b. Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, 1797–1806.

W. H. Cooper, R. B. Gallupe, S. Pollard, and J. Cadsby. 1998. Some liberating effects of anonymous electronic brainstorming. *Small Group Research* 29, 2, 147–178.

P. T. Costa and R. R. MacCrae. 1992. *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO FFI): professional manual*. Psychological Assessment Resources.

A. J. Cropley. 2000. Defining and measuring creativity: Are creativity tests worth using? *Roeper Review* 23, 2, 72–79.

M. Crossley, N. J. Davies, A. J. McGrath, and M. A. Z. Rejman-Greene. 1998. The knowledge garden. *BT Technology Journal* 17, 1, 76–84.

M. Csikszentmihalyi. 1990. *Creativity: Flow and the psychology of discovery and invention.* HarperCollins Publishers.

G. Cumming. 2013. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

E. De Bono. 1985. *Six Thinking Hats*. Key Porter Books.

K. E. De Stobbeleir, S. J. Ashford, and D. Buyens. 2011. Self-regulation of creativity at work: The role of feedback-seeking behavior in creative performance. *Academy of Management Journal* 54, 4, 811–831.

D. L. Dean, J. M. Hender, T. L. Rodgers, and E. Santanen. 2006. Identifying good ideas: constructs and scales for idea evaluation. *Journal of Association for Information Systems* 7, 10, 646–699.

E. L. Deci and R. M. Ryan. 1975. *Intrinsic motivation*. Wiley Online Library.

J. Dehler, D. Bodemer, J. Buder, and F. W. Hesse. 2011. Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior* 27, 3, 1068–1078.

A. R. Dennis and B. A. Reinicke. 2004. Beta versus VHS and the acceptance of electronic brainstorming technology. *MIS Quarterly* 28, 1, 1–20.

A. R. Dennis and J. S. Valacich. 1993. Computer brainstorms: More heads are better than one. *Journal of applied psychology* 78, 4, 531–537.

D. M. DeRosa, C. L. Smith, and D. A. Hantula. 2007. The medium matters: Mining the long-promised merit of group interaction in creative idea generation tasks in a meta-analysis of the electronic group brainstorming literature. *Computers in Human Behavior* 23, 3, 1549–1581.

S. Deterding, D. Dixon, R. Khaled, and L. Nacke. 2011. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11)*. ACM, 9–15.

M. Deutsch. 1949. An experimental study of the effects of cooperation and competition upon group process. *Human relations* 2, 3, 199–231.

M. Diehl and W. Stroebe. 1987. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of personality and social psychology* 53, 3, 497–509.

J. M. Digman and J. Inouye. 1986. Further specification of the five robust factors of personality. *Journal of Personality and Social Psychology* 50, 1, 116–123.

P. Dillenbourg and P. Jermann. 2007. Designing integrative scripts. In *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives*. Springer, 275–301.

R. Dilts. 1995. *Strategies of genius.* Meta Publications.

J. M. DiMicco. 2004. Designing interfaces that influence group processes. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04)*. ACM, 1041–1042.

J. M. DiMicco and W. Bender. 2004. Second messenger: Increasing the visibility of minority viewpoints with a face-to-face collaboration tool. In *Proceedings of the 9th International Conference on Intelligent User Interfaces (IUI '04)*. ACM, 232–234.

J. M. DiMicco and W. Bender. 2007. Group reactions to visual feedback tools. In *Proceedings of the 2nd International Conference on Persuasive Technology (PERSUASIVE '07)*. Springer, 132–143.

J. M. DiMicco and K. J. Hollenbach. 2006. Visualization of Audio: A social tool for face-to-face groups. In *Social Visualization Workshop, Conference on Human Factors in Computing Systems (CHI '06)*.

J. M. DiMicco, K. J. Hollenbach, and W. Bender. 2006. Using visualizations to review a group's interaction dynamics. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, 706–711.

J. M. DiMicco, K. J. Hollenbach, A. Pandolfo, and W. Bender. 2007. The impact of increased awareness while face-to-face. *Human–Computer Interaction* 22, 1-2, 47–96.

J. M. DiMicco, A. Pandolfo, and W. Bender. 2004. Influencing group participation with a shared display. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, 614–623.

J. Donath. 2002. A semantic approach to visualizing online conversations. *Communications of the ACM - Supporting community and building social capital* 45, 4, 45–49.

J. Donath, K. Karahalios, and F. Viégas. 2000. Visiphone. *Proceedings of ICAD 2000*.

J. Donath and F. B. Viégas. 2002. The Chat Circles series: Explorations in designing abstract graphical communication interfaces. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '02)*. ACM, 359–369.

P. Dourish and V. Bellotti. 1992. Awareness and coordination in shared workspaces. *Proceedings of the 1992 ACM conference on Computer-supported cooperative work - CSCW '92* November, 107–114.

P. Dourish and S. Bly. 1992. Portholes: Supporting awareness in a distributed work group. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, 541–547.

M. Doyle and D. Straus. 1976. *How to make meetings work.* Jove Books New York.

P. Dragicevic. 2015. Hci statistics without p-values. *[Research Report] RR-8738*, 32.

V. J. Dubrovsky, S. Kiesler, and B. N. Sethna. 1991. The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups. *Human–Computer Interaction* 6, 2, 119–146.

R. J. Dufresne, W. J. Gerace, W. J. Leonard, J. P. Mestre, and L. Wenk. 1996. Classtalk: A classroom communication system for active learning. *Journal of computing in higher education* 7, 2, 3–47.

P. C. Earley. 1988. Computer-generated performance feedback in the magazine-subscription industry. *Organizational Behavior and Human Decision Processes* 41, 1, 50 – 64.

M. Ebner, C. Haintz, K. Pichler, and S. Schön. 2014. Technologiegestützte Echtzeit-Interaktion in Massenvorlesungen im Hörsaal. Entwicklung und Erprobung eines digitalen Backchannels während der Vorlesung. In *Lernräume gestalten-Bildungskontexte vielfältig denken*. 567–578. In German.

B. Eggen and K. Van Mensvoort. 2009. Making sense of what is going on ‘around’: Designing environmental awareness information displays. In *Awareness Systems: Advances in Theory, Methodology and Design*. Springer, 99–124.

C. A. Ellis, S. J. Gibbs, and G. Rein. 1991. Groupware: Some issues and experiences. *Commun. ACM* 34, 1, 39–58.

M. R. Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1, 32–64.

D. Epstein, F. Cordeiro, E. Bales, J. Fogarty, and S. Munson. 2014. Taming data complexity in lifelogs: Exploring visual cuts of personal informatics data. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS ’14)*. ACM, 667–676.

F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong. 2016. The geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing* 7, 2, 190–202.

F. Eyben, F. Weninger, F. Gross, and B. Schuller. 2013. Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia (MM ’13)*. ACM, 835–838.

U. Farooq, J. M. Carroll, and C. H. Ganoe. 2005. Supporting creativity in distributed scientific communities. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (GROUP ’05)*. ACM, 217–226.

G. Fischer. 2004. Social creativity: Turning barriers into opportunities for collaborative design. In *Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices (PDC ’04)*. ACM, 152–161.

D. W. Fiske. 1949. Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology* 44, 3, 329–344.

B. Flinn and H. Maurer. 1996. *Levels of anonymity*. Springer.

A. Freeley and D. Steinberg. 2013. *Argumentation and debate*. Cengage Learning.

M. Fuge, J. Stroud, and A. Agogino. 2013. Automatically inferring metrics for design creativity. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (ASME ’13)*. American Society of Mechanical Engineers, 10.

M. Gagné and E. L. Deci. 2005. Self-determination theory and work motivation. *Journal of Organizational Behavior* 26, 4, 331–362.

R. B. Gallupe, L. M. Bastianutti, and W. H. Cooper. 1991. Unblocking brainstorms. *Journal of Applied Psychology* 76, 1, 137.

R. G. Geen. 1991. Social motivation. *Annual review of psychology* 42, 1, 377–399.

V. Gehlen–Baum, A. Pohl, A. Weinberger, and F. Bry. 2012. Backstage–designing a backchannel for large lectures. In *Proceedings of the European Conference on Technology Enhanced Learning (EC-TEL'2012)*. Springer, 459–464.

V. Gehlen–Baum, A. Weinberger, A. Pohl, and F. Bry. 2014. Technology use in lectures to enhance students' attention. In *Open Learning and Teaching in Educational Communities: Proceedings of the 9th European Conference on Technology Enhanced Learning, (EC-TEL '14)*. Springer, 125–137.

J. M. George and J. Zhou. 2007. Dual tuning in a supportive context: Joint contributions of positive mood, negative mood, and supervisory behaviors to employee creativity. *Academy of Management Journal* 50, 3, 605–622.

B. Gerhart and M. Fang. 2015. Pay, intrinsic motivation, extrinsic motivation, performance, and creativity in the workplace: Revisiting long-held beliefs. *Annual Review of Organizational Psychology and Organizational Behavior* 2, 1, 489–521.

H. Geschka. 1978. Introduction and use of idea-generating methods. *Research management* 21, 3, 25–28.

F. Geyer, J. Budzinski, and H. Reiterer. 2012. Ideavis: A hybrid workspace and interactive visualization for paper-based collaborative sketching sessions. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12)*. ACM, 331–340.

R. Glaser. 1963. Instructional technology and the measurement of learing outcomes: Some questions. *American psychologist* 18, 8, 519–521.

R. Glaser and A. J. Nitko. 1970. *Measurement in Learning and Instruction.* Technical Report.

B. Goodman, F. Linton, R. Gaimari, J. Hitzeman, H. Ross, and G. Zarrella. 2005. Using dialogue features to predict trouble during collaborative learning. *User Modeling and User–Adapted Interaction* 15, 1-2, 85–134.

H. G. Gough. 1960. The adjective check list as a personality assessment research technique. *Psychological Reports* 6, 1, 107–122.

T. R. Gruber. 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition* 5, 2, 199 – 220.

J. Guilford. 1983. Transformotion abilities or functions. *The Journal of creative behavior* 17, 2, 75–83.

J. P. Guilford. 1950. Creativity. *American Psychologist* 5, 9, 444 – 454.

F. Guimbretière, M. Stone, and T. Winograd. 2001. Fluid interaction with high-resolution wall-size displays. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST '01)*. ACM, 21–30.

C. Gutwin and S. Greenberg. 1996. Workspace awareness for groupware. In *Conference Companion on Human Factors in Computing Systems (CHI '96)*. ACM, 208–209.

C. Gutwin and S. Greenberg. 1998. Effects of awareness support on groupware usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., 511–518.

C. Gutwin and S. Greenberg. 2000. The mechanics of collaboration: Developing low cost usability evaluation methods for shared workspaces. In *Proceedings of the 9th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE '00)*. IEEE Computer Society, 98–103.

C. Gutwin and S. Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)* 11, 3, 411–446.

J. R. Hackman and G. R. Oldham. 1980. *Work redesign*. Addison-Wesley.

J. Hailpern, E. Hinterbichler, C. Leppert, D. Cook, and B. P. Bailey. 2007. TEAM STORM: Demonstrating an interaction model for working with multiple ideas during creative group work. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity &Amp; Cognition (C&C '07)*. ACM, 193–202.

R. K. Hambleton and M. R. Novick. 1973. Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 159–170.

D. Harry, J. Green, and J. Donath. 2009. Backchan.nl: Integrating backchannels in physical space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, 1361–1370.

J. Hattie. 1992. *Self-concept*. Lawrence Erlbaum Associates, Inc.

J. Hattie and H. Timperley. 2007. The power of feedback. *Review of Educational Research* 77, 1, 81–112.

D. Hausen, S. Boring, C. Lueling, S. Rodestock, and A. Butz. 2012. StaTube: Facilitating state management in instant messaging systems. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction (TEI '12)*. ACM, 283–290.

B. A. Hennessey and T. M. Amabile. 2010. Creativity. *Annual Review of Psychology* 61, 1, 569–598.

O. Hilliges, L. Terrenghi, S. Boring, D. Kim, H. Richter, and A. Butz. 2007. Designing for collaborative creative problem solving. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity &Amp; Cognition (C&C '07)*. ACM, 137–146.

J. B. Hirschberg and A. Rosenberg. 2005. Acoustic, prosodic and lexical correlates of charismatic speech. In *Ninth European Conference on Speech Communication and Technology (INTERSPEECH '05)*.

E. Howie, S. Sy, L. Ford, and K. J. Vicente. 2000. Human-computer interface design can reduce misperceptions of feedback. *System Dynamics Review* 16, 3, 151–171.

B. Huber, S. Tausch, and H. Hußmann. 2014. Supporting debates with a real-time feedback system. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems (CHI EA '14)*. ACM, 2257–2262.

S. Hunter, P. Maes, S. Scott, and H. Kaufman. 2011. MemTable: An integrated system for capture and recall of shared histories in group workspaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, 3305–3314.

K. Inkpen, K. Hawkey, M. Kellar, R. Mandryk, K. Parker, D. Reilly, S. Scott, and T. Whalen. 2005. Exploring display factors that influence co-located collaboration: Angle, size, number, and user arrangement. *Proceedings of HCI International*.

A. Inselberg and B. Dimsdale. 1991. Parallel coordinates. In *Human-Machine Interactive Systems*. Springer, 199–233.

International Debate Education Association. 2004. *The debatabase book: a must-have guide for successful debate*. Idea Press Books.

H. Ishii, M. Kobayashi, and J. Grudin. 1993. Integration of interpersonal space and shared workspace: ClearBoard design and experiments. *ACM Transactions on Information Systems (TOIS)* 11, 4, 349–375.

H. Ishii, C. Wisneski, S. Brave, A. Dahley, M. Gorbet, B. Ullmer, and P. Yarin. 1998. ambientROOM: Integrating ambient media with architectural space. In *CHI 98 Conference Summary on Human Factors in Computing Systems (CHI '98)*. ACM, 173–174.

J. Israel and R. Aiken. 2007. Supporting collaborative learning with an intelligent web–based system. *International Journal of Artificial Intelligence in Education* 17, 1, 3–40.

A. Jaco, S. Buisine, J. Barré, A. Aoussat, and F. Vernier. 2014. Trains of thought on the tabletop: Visualizing association of ideas improves creativity. *Personal Ubiquitous Computing* 18, 5, 1159–1167.

J. Janssen and D. Bodemer. 2013. Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist* 48, 1, 40–55.

P. Jermann and P. Dillenbourg. 2008. Group mirrors to support interaction regulation in collaborative problem solving. *Computers & Education* 51, 1, 279–296.

P. Jermann, A. Soller, and M. Muehlenbrock. 2001. From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. In *European Conference on Computer-Supported Collaborative Learning (EuroCSCL '01)*. 324–331.

R. Johansen. 1988. *GroupWare: Computer Support for Business Teams*. The Free Press.

D. W. Johnson and R. T. Johnson. 1989. *Cooperation and competition: Theory and research.* Interaction Book Company.

D. W. Johnson and R. T. Johnson. 2005. New developments in social interdependence theory. *Genetic, social, and general psychology monographs* 131, 4, 285–358.

D. W. Johnson and R. T. Johnson. 2011. Intellectual legacy: Cooperation and competition. In *Conflict, Interdependence, and Justice*. Springer, 41–63.

D. W. Johnson, R. T. Johnson, and L. Scott. 1978. The effects of cooperative and individualized instruction on student attitudes and achievement. *The Journal of Social Psychology* 104, 2, 207–216.

D. H. Jonassen and B. Kim. 2010. Arguing to learn and learning to argue: design justifications and guidelines. *Educational Technology Research and Development* 58, 4, 439–457.

B. Kaltsounts and L. Honeywell. 1980. Additional instruments useful in studying creative behavior and creative talent. *The Journal of Creative Behavior* 14, 1, 56–67.

M. Kam, J. Wang, A. Iles, E. Tse, J. Chiu, D. Glaser, O. Tarshish, and J. Canny. 2005. Livenotes: A system for cooperative and augmented note-taking in lectures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05).* ACM, 531–540.

S. A. Karabenick and J. R. Knapp. 1988. Effects of computer privacy on help-seeking. *Journal of Applied Social Psychology* 18, 6, 461–472.

N. Karacapilidis and D. Papadias. 2001. Computer supported argumentation and collaborative decision making: the HERMES system. *Information Systems* 26, 4, 259 – 277.

K. Karahalios. 2009. Social mirrors as social signals: Transforming audio into graphics. *IEEE Computer Graphics and Applications* 29, 5, 22–32.

K. Karahalios and T. Bergstrom. 2006. Visualizing audio in group table conversation. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems, 2006. TableTop 2006.* IEEE, 131–134.

S. J. Karau and K. D. Williams. 1993. Social loafing: A meta-analytic review and theoretical integration. *Interpersonal Relations and Group Processes* 65, 4, 681–706.

J. F. Kelley. 1983. An empirical methodology for writing user-friendly natural language computer applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '83)*. ACM, 193–196.

B. Kerr and C. Gagliardi. 2003. Measuring creativity in research and practice. In *Positive Psychological Assessment: A Handbook of Models and Measures*. American Psychological Association, 155–169.

T. Kim, A. Chang, L. Holland, and A. S. Pentland. 2008. Meeting Mediator: Enhancing group collaboration using sociometric feedback. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work (CSCW '08)*. ACM, 457–466.

T. Kim and A. S. Pentland. 2009. *Understanding Effects of Feedback on Group Collaboration*. Technical Report. AAAI Press.

J. W. Kinch. 1963. A formalized theory of the self-concept. *American Journal of Sociology* 68, 4, 481–486.

J. W. Kinch. 1968. Experiments on factors related to self-concept change. *The Journal of Social Psychology* 74, 2, 251–258.

M. Kirsch-Pinheiro, J. V. de Lima, and M. R. S. Borges. 2003. A framework for awareness support in groupware systems. *Computers in Industry* 52, 1, 47–57.

M. Kirsch-Pinheiro, J. Lima, and M. Borges. 2001. Awareness em sistemas de groupware. *4 Jornadas Iberoamericanas de Ingenieria de Requisitos y Ambientes de Software*, 68. In Spanish.

M. Klein and L. Iandoli. 2008. Supporting collaborative deliberation using a large-scale argumentation system: the MIT Collaboratorium. *The MIT Center for Digital Business*.

S. R. Klemmer, M. W. Newman, R. Farrell, M. Bilezikjian, and J. A. Landay. 2001. The designers' outpost: A tangible interface for collaborative web site. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST '01)*. ACM, 1–10.

A. N. Kluger and S. Adler. 1993. Person- versus computer-mediated feedback. *Computers in Human Behavior* 9, 1, 1 – 16.

A. N. Kluger and A. DeNisi. 1996. The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin* 199, 2, 254–284.

I. Kosan. 2013. *Erweiterung und Evaluation eines Group Mirror Frameworks zur Unterstützung von Brainstorming*. Bachelor thesis. Ludwig-Maximilians-Universität München.

R. Kosara, F. Bendix, and H. Hauser. 2006. Parallel sets: interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics* 12, 4, 558–568.

T. Koschmann. 2003. CSCL, argumentation, and deweyan inquiry. In *Arguing to learn*. Springer, 261–269.

D. Kuhn. 2001. How do people know? *Psychological Science* 12, 1, 1–8.

D. Kuhn and W. Goh. 2005. Arguing on the computer. In *Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!* International Society of the Learning Sciences, 346–352.

D. Kuhn, V. Shaw, and M. Felton. 1997. Effects of dyadic interaction on argumentive reasoning. *Cognition and Instruction* 15, 3, 287–315.

J. A. Kulik and C.-L. C. Kulik. 1988. Timing of feedback and verbal learning. *Review of Educational Research* 58, 1, 79–97.

O. Kulyk, J. Wang, and J. Terken. 2006. Real-time feedback on nonverbal behaviour to enhance social dynamics in small group meetings. In *Machine Learning for Multimodal Interaction: Second International Workshop, MLMI 2005, Edinburgh, UK, July 11-13, 2005, Revised Selected Papers*. Springer, 150–161.

K. Kurihara, M. Goto, J. Ogata, Y. Matsusaka, and T. Igarashi. 2007. Presentation Sensei: a presentation training system using speech and image processing. In *Proceedings of the 9th international conference on Multimodal interfaces*. ACM, 358–365.

H. Lamm and G. Trommsdorff. 1973. Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review. *European journal of social psychology* 3, 4, 361–388.

M. Lantin and G. Judelman. 2006. Flowergarden: An interactive visualization of concept-sharing. In *Proceedings of the CHI'06 Workshop on About Face: Interface Creative Engagement in New Media Arts and HCI.*

G. P. Latham and E. A. Locke. 1991. Self-regulation through goal setting. *Organizational behavior and human decision processes* 50, 2, 212–247.

R. S. Lazarus. 1991. *Emotion and adaptation.* Oxford University Press.

S. Leitão. 2000. The Potential of Argument in Knowledge Building. *Human Development* 43, 6, 332–360.

R. Lengler and M. J. Eppler. 2007. Towards a periodic table of visualization methods of management. In *Proceedings of the IASTED International Conference on Graphics and Visualization in Engineering (GVE '07).* ACTA Press, 83–88.

G. Leshed, J. T. Hancock, D. Cosley, P. L. McLeod, and G. Gay. 2007. Feedback for guiding reflection on teamwork practices. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work (GROUP '07)*. ACM, 217–220.

G. Leshed, D. Perez, J. T. Hancock, D. Cosley, J. Birnholtz, S. Lee, P. L. McLeod, and G. Gay. 2009. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, 537–546.

M. C. Linn, P. Bell, and S. Hsi. 1998. Using the internet to enhance student understanding of science:the knowledge integration environment. *Interactive Learning Environments* 6, 1-2, 4–38.

G. L. Lohse, K. Biolsi, N. Walker, and H. H. Rueter. 1994. A classification of visual representations. *Communications of the ACM* 37, 12, 36–49.

F. Loll and N. Pinkwart. 2009. Using collaborative filtering algorithms as elearning tools. In *Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS '09)*. IEEE, 10.

X. Lu. 1998. *Rhetoric in ancient China, fifth to third century, BCE: A comparison with classical Greek rhetoric*. University of South Carolina Press.

R. Lyster and L. Ranta. 1997. Corrective feedback and learner uptake negotiation of form in communication classrooms. *Studies in Second Language Acquisition* 20, 37–66.

K. R. MacCrimmon and C. Wagner. 1994. Stimulating ideas through creative software. *Management Science* 40, 11, 1514–1532.

M. L. Maher. 2010. Evaluating creativity in humans, computers, and collectively intelligent systems. In *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design (DESIRE '10)*. Desire Network, 22–28.

R. M. Maldonado, J. Kay, K. Yacef, and B. Schwendimann. 2012. An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *Intelligent Tutoring Systems*. Springer, 482–492.

J. B. Maller. 1972. *Coöperation and competition: an experimental study in motivation*. Bureau of Publications, Teachers College, Columbia University.

R. L. Mandryk, S. D. Scott, and K. M. Inkpen. 2002. Display factors influencing co-located collaboration. In *Interactive Poster at ACM Conference on Computer-Supported Cooperative Work (CSCW '02)*.

J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. 2003. Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, 169–176.

G. Mark, L. Fuchs, and M. Sohlenkamp. 1997. Supporting groupware conventions through contextual awareness. In *Proceedings of The Fifth European Conference on Computer Supported Cooperative Work*. Springer, 253–268.

P. Markopoulos and W. Mackay. 2009. *Awareness systems: Advances in theory, methodology and design*. Springer.

R. Martinez, J. Kay, J. R. Wallace, and K. Yacef. 2011. Modelling symmetry of activity as an indicator of collocated group collaboration. In *User Modeling, Adaption and Personalization*. Springer, 207–218.

L. Mason and M. Santi. 1994. *Argumentation Structure and Metacognition in Constructing Shared Knowledge at School*. Technical Report. Paper presented at the Annual Meeting of the American Educational Research Association.

P. Mathur and K. Karahalios. 2009. Visualizing remote voice conversations. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)*. ACM, 4675–4680.

K. R. Max Wertheimer. 1944. Gestalt theory. *Social Research* 11, 1, 78–99.

M. A. May and L. W. Doob. 1937. *Competition and Cooperation*. Vol. 25. Social Science Research Council Bulletin.

S. McAlister, A. Revenscroft, and E. Scanlon. 2004. Combining interaction and context design to support collaborative argumentation using a tool for synchronous CMC. *Journal of Computer Assisted Learning* 20, 3, 194 – 204.

S. E. McDaniel and T. Brinck. 1997. Awareness in collaborative systems. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems (CHI EA '97)*. ACM, 237–237.

M. Mead. 1937. *Co-operation and competition among primitive peoples*. McGraw-Hill Publications in Sociology.

K. L. Mills. 2003. Computer-supported cooperative work. In *Encyclopedia of Library and Information Science*. Marcel Dekker, 666–677.

J. Mu, K. Stegmann, E. Mayfield, C. Rosé, and F. Fischer. 2012. The ACODEA framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International Journal of Computer-Supported Collaborative Learning* 7, 2, 285–305.

B. Mullen, C. Johnson, and E. Salas. 1991. Productivity loss in brainstorming groups: A meta-analytic integration. *Basic and applied social psychology* 12, 1, 3–23.

J. Nakahara, S. Hisamatsu, K. Yaegashi, and Y. Yamauchi. 2005. iTree: Does the Mobile Phone Encourage Learners to Be More Involved in Collaborative Learning?. In *Proceedings of the 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: The Next 10 Years! (CSCL '05)*. International Society of the Learning Sciences, 470–478.

D. M. Nebeker and B. C. Tatum. 1993. The effects of computer monitoring, standards, and rewards on work performance, job satisfaction, and stress. *Journal of Applied Social Psychology* 23, 7, 508–536.

G. Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education* 15, 5, 625–632.

J. F. Nunamaker, R. O. Briggs, D. D. Mittleman, D. R. Vogel, and B. A. Pierre. 1996. Lessons from a dozen years of group support systems research: A discussion of lab and field findings. *Journal of Management Information Systems* 13, 3, 163–207.

J. F. Nunamaker, A. R. Dennis, J. S. Valacich, D. Vogel, and J. F. George. 1991. Electronic meeting systems. *Commun. ACM* 34, 7, 40–61.

F. Nußberger. 2014. *Design and evaluation of a prototype for co-located collaborative meetings using role-based creativity strategies*. Bachelor thesis. Ludwig-Maximilians-Universität München.

K. Ogawa, Y. Hori, T. Takeuchi, T. Narumi, T. Tanikawa, and M. Hirose. 2012. Table talk enhancer: A tabletop system for enhancing and balancing mealtime conversations using utterance rates. In *Proceedings of the ACM Multimedia 2012 Workshop on Multimedia for Cooking and Eating Activities (CEA '12)*. ACM, 25–30.

A. F. Osborn. 1953. *Applied Imagination: Principles and Procedures of Creative Thinking*. Scribner.

N. L. Oxley, M. T. Dzindolet, and P. B. Paulus. 1996. The effects of facilitators on the performance of brainstorming groups. *Journal of Social Behavior and Personality* 11, 4, 633.

P. B. Paulus and M. T. Dzindolet. 1993. Social influence processes in group brainstorming. *Journal of Personality and Social Psychology* 64, 4, 575–586.

P. B. Paulus and B. A. Nijstad. 2003. *Group creativity: Innovation through collaboration*. Oxford University Press.

P. B. Paulus and H.-C. Yang. 2000. Idea generation in groups: A basis for creativity in organizations. *Organizational behavior and human decision processes* 82, 1, 76–87.

F. Perteneder, C. Grossauer, T. Seifried, J. Walney, J. Brosz, A. Tang, S. Carpendale, and M. Haller. 2012. Idea playground: When brainstorming is not enough. In *Proceedings of the workshop Blended Interaction in Conjunction with the Conference on Advanced visual interfaces (AVI '12)*.

C. Phielix, F. J. Prins, and P. A. Kirschner. 2010. Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior* 26, 2, 151–161.

C. Phielix, F. J. Prins, P. A. Kirschner, G. Erkens, and J. Jaspers. 2011. Group awareness of social and cognitive performance in a CSCL environment: effects of a peer feedback and reflection tool. *Computers in Human Behavior* 27, 3, 1087–1102.

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. 2015. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-120.

N. Pinkwart, V. Aleven, K. Ashley, and C. Lynch. 2006. Toward legal argument instruction with graph grammars and collaborative filtering techniques. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems (ITS '06)*. Springer, 227–236.

N. Pinkwart, V. Aleven, K. Ashley, and C. Lynch. 2007. Evaluating legal argument instruction with graphical representations using LARGO. In *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AI-ED 2007)*, Vol. 158. IOS Press, 101–108.

N. Pinkwart, C. Lynch, K. Ashley, and V. Aleven. 2008. Re-evaluating LARGO in the classroom: Are diagrams better than text for teaching argumentation skills? In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS '08)*. Springer, 90–100.

T. S. Pittman, M. E. Davey, K. A. Alafat, K. V. Wetherill, and N. A. Kramer. 1980. Informational versus controlling verbal rewards. *Personality and Social Psychology Bulletin* 6, 2, 228–233.

J. A. Plucker, M. Qian, and S. Wang. 2011. Is originality in the eye of the beholder? Comparison of scoring techniques in the assessment of divergent thinking. *The Journal of Creative Behavior* 45, 1, 1–22.

J. A. Plucker and M. A. Runco. 1998. The death of creativity measurement has been greatly exaggerated: Current issues, recent advances, and future directions in creativity assessment. *Roeper Review* 21, 1, 36–39.

A. Pohl, F. Bry, J. Schwarz, and M. Gottstein. 2012. Sensing the classroom: Improving awareness and self-awareness of students in backstage. In *15th International Conference on Interactive Collaborative Learning (ICL)*. 8.

A. Pohl, V. Gehlen–Baum, and F. Bry. 2011. Introducing backstage – a digital backchannel for large class lectures. *Interactive Technology and Smart Education* 8, 3, 186–200.

A. Pohl, V. Gehlen–Baum, and F. Bry. 2012. Enhancing the digital backchannel backstage on the basis of a formative user study. *International Journal of Emerging Technologies in Learning (iJET)* 7, 1, 33–41.

J. Polleti, D. Baur, T. Tang, and S. Carpendale. 2012. ECO|Balance – Exploring Design Issues for Mobile Persuasion. *Personal Informatics in Practice: Improving Quality of Life Through Data Workshop at the conference on Human factors in computing systems (CHI '12).*

Z. Pousman and J. Stasko. 2006. A taxonomy of ambient information systems: four patterns of design. *Proceedings of the Working Conference on Advanced Visual Interfaces*, 67–74.

A. Raltchev. 2013. *Comparing Horizontal with Vertical Displays for a Feedback System*. Project thesis. Ludwig-Maximilians-Universität München.

M. Ratto, R. B. Shapiro, T. M. Truong, and W. G. Griswold. 2003. The activeclass project: Experiments in encouraging classroom participation. In *Designing for Change in Networked Learning Environments*. Springer, 477–486.

W. Reinhardt, M. Sievers, J. Magenheim, D. Kundisch, P. Herrmann, M. Beutner, and A. Zoyke. 2012. Pingo: peer instruction for very large groups. In *21st Century Learning for 21st Century Skills*. Springer, 507–512.

C. Reithmeier. 2013. *Evaluation of different feedback types for co-located collaboration*. Bachelor thesis. Ludwig-Maximilians-Universität München.

R. D. Rieke, M. O. Sillars, and T. R. Peterson. 2009. *Argumentation and critical decision making*. Pearson, Allyn and Bacon.

T. Rindlbacher. 2015. *Evaluating different complexity levels of real-time feedback during argumentation*. Bachelor thesis. Ludwig-Maximilians-Universität München.

M. Rittenbruch and G. McEwan. 2009. An historical reflection of awareness in collaboration. In *Awareness Systems*. Springer, 3–48.

T. Rodden. 1996. Populating the application: A model of awareness for cooperative applications. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW '96)*. ACM, 87–96.

R. Rodenstein and J. S. Donath. 2000. Talking in circles: Designing a spatially-grounded audioconferencing environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '00)*. ACM, 81–88.

Y. Rogers. 2011. Interaction design gone wild: Striving for wild theory. *interactions* 18, 4, 58–62.

Y. Rogers and S. Lindley. 2004. Collaborating around vertical and horizontal large interactive displays: which way is best? *Interacting with Computers* 16, 6, 1133–1152.

J. Rooksby, M. Rost, A. Morrison, and M. C. Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, 1163–1172.

M. C. Roy, S. Gauvin, and M. Limayem. 1996. Electronic Group Brainstorming: The Role of Feedback on Productivity. *Small Group Research* 27, 2, 215–247.

M. A. Runco. 1997. *The creativity research handbook*. Vol. 1. Hampton Press.

M. A. Runco and S. R. Pritzker. 1999. *Encyclopedia of creativity, Vol. 1 A–H and Vol. 2 I–Z with indexes.* Academic Press.

R. M. Ryan and E. L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1, 68.

K. Sachmann. 2014. *Vergleich von verschiedenen Displayumgebungen für Group Mirror Systeme*. Bachelor thesis. Ludwig-Maximilians-Universität München.

B. Saket, S. Yang, H. Tan, K. Yatani, and D. Edge. 2014. Talkzones: section-based time support for presentations. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 263–272.

T. Sather. 1999. *Pros and cons: a debater's handbook*. Taylor & Francis.

P. K. Schank. 1995. *Computational tools for modeling and aiding reasoning: Assessing and applying the theory of explanatory coherence*. Ph.D. Dissertation. University of California at Berkeley.

O. Scheuer, F. Loll, N. Pinkwart, and B. M. McLaren. 2010. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning* 5, 1, 43–102.

G. Schiavo, A. Cappelletti, E. Mencarini, O. Stock, and M. Zancanaro. 2014. Overt or subtle? supporting group conversations with automatically targeted directives. In *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI '14)*. ACM, 225–234.

G. Schiavo, A. Cappelletti, E. Mencarini, O. Stock, and M. Zancanaro. 2016. Influencing participation in group brainstorming through ambient intelligence. *International Journal of Human-Computer Interaction* 32, 3.

K. Schmidt. 2002. The problem with awareness: Introductory remarks on awareness in CSCW. *Computer Supported Cooperative Work (CSCW)* 11, 3-4, 285–298.

D. H. Schunk. 1990. Goal setting and self-efficacy during self-regulated learning. *Educational psychologist* 25, 1, 71–86.

B. B. Schwarz and A. Glassner. 2007. The role of floor control and of ontology in argumentative activities with discussion-based tools. *International Journal of Computer-Supported Collaborative Learning* 2, 4, 449–478.

C. W. Sherif. 1976. The social context of competition. *Joy and sadness in children's sports*, 81–97.

B. Shneiderman. 1996. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*. IEEE, 336–343.

B. Shneiderman. 2000. Creating creativity: User interfaces for supporting innovation. *ACM Transactions of Computer-Human Interaction* 7, 1, 114–138.

V. J. Shute. 2008. Focus on formative feedback. *Review of educational research* 78, 1, 153–189.

C.-L. Sia, B. C. Tan, and K.-K. Wei. 2002. Group polarization and computer-mediated communication: Effects of communication cues, social presence, and anonymity. *Information Systems Research* 13, 1, 70–90.

P. J. Silvia, B. P. Winterstein, J. T. Willse, C. M. Barona, J. T. Cram, K. I. Hess, J. L. Martinez, and C. A. Richard. 2008. Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts* 2, 2, 68.

B. Simon, K. Davis, W. G. Griswold, M. Kelly, and R. Malani. 2008. Noteblogging: Taking note taking public. In *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education (SIGCSE '08)*. ACM, 417–421.

D. J. Simons and D. T. Levin. 1997. Change blindness. *Trends in Cognitive Sciences* 1, 7, 261 – 267.

J. J. Sosik, B. J. Avolio, and S. S. Kahai. 1998. Inspiring group creativity: Comparing anonymous and identified electronic brainstorming. *Small Group Research* 29, 1, 3–31.

K. Stegmann, C. Wecker, A. Weinberger, and F. Fischer. 2012. Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. *Instructional Science* 40, 2, 297–323.

K. Stegmann, A. Weinberger, and F. Fischer. 2007. Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning* 2, 4, 421–447.

F. Steinberger. 2013. *Entwurf und Evaluierung eines Group Mirrors zur Untersttützung der Walt Disney Methode*. Practical course. Ludwig-Maximilians-Universität München.

R. J. Sternberg. 1999. *Handbook of creativity*. Cambridge University Press.

N. A. Streitz, J. Geißler, T. Holmer, S. Konomi, C. Müller-Tomfelde, W. Reischl, P. Rexroth, P. Seitz, and R. Steinmetz. 1999. i-LAND: An interactive landscape for creativity and innovation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, 120–127.

S. Streng. 2010. Supporting argumentative knowledge construction in face-to-face settings: From ArgueTable to ArgueWall. In *Proceedings of the Conference on Computer-Supported Collaborative Learning*. International Society of the Learning Sciences, 716–720.

S. Streng. 2012. *The Role of personal and shared displays in scripted collaborative learning*. Ph.D. Dissertation. LMU Münnchen.

S. Streng, K. Stegmann, H. Hußmann, and F. Fischer. 2009. Metaphor or diagram?: Comparing different representations for group mirrors. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7 (OZCHI '09)*. ACM, 249–256.

R. Stults. 1986. *Media space*. Technical Report.

J. Sturm, O. H.-v. Herwijnen, A. Eyck, and J. Terken. 2007. Influencing social dynamics in meetings through a peripheral display. In *Proceedings of the 9th International Conference on Multimodal Interfaces (ICMI '07)*. ACM, 263–270.

J. Sturm, R. Iqbal, O. Kulyk, J. Wang, and J. Terken. 2005. Peripheral feedback on participation level to support meetings and lectures. In *Proceeding of Conference on Designing Pleasurable Products and Interfaces (DPPI)*. Eindhoven Technical University Press.

J. Sturm, R. Iqbal, and J. Terken. 2006. Development of peripheral feedback to support lectures. In *Machine Learning for Multimodal Interaction*. Springer, 138–149.

D. D. Suthers. 2003. Representational guidance for collaborative inquiry. In *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments*. Springer, 27–46.

D. D. Suthers, J. Connelly, A. Lesgold, M. Paolucci, E. E. Toth, J. Toth, and A. Weiner. 2001. Representational and advisory guidance for students learning scientific inquiry. In *Smart Machines in Education*. MIT Press, 7–35.

W. B. Swann, B. W. Pelham, and T. R. Chidester. 1988. Change through paradox: using self-verification to alter beliefs. *Journal of personality and social psychology* 54, 2, 268–273.

S. Ta. 2014. *Verbundenheit oder Wettstreit? Vergleich verschiedener Visualisierungen Unterstützung von Brainstorming*. Bachelor thesis. Ludwig-Maximilians-Universität München.

D. Tam, K. E. MacLean, J. McGrenere, and K. J. Kuchenbecker. 2013. The design and field observation of a haptic notification system for timing awareness during oral presentations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1689–1698.

D. Tannen. 1998. *The argument culture: Moving from debate to dialogue*. Ballantine Books.

S. Tausch. 2011. *Untersuchung des Einflusses von Displaykonfigurationen auf phasenorientiertes Gruppenlernen*. Diploma thesis. Ludwig-Maximilians-Universität München.

S. Tausch, D. Hausen, I. Kosan, A. Raltchev, and H. Hussmann. 2014. Groupgarden: Supporting brainstorming through a metaphorical group mirror on table or wall. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*. ACM, 541–550.

S. Tausch, F. Nußberger, and H. Hußmann. 2015a. Supporting the disney method with an interactive feedback system. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, 1013–1018.

S. Tausch, F. Steinberger, and H. Hußmann. 2015b. Thinking like disney: Supporting the disney method using ambient feedback based on group performance. In *Proceedings of the 15th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT '15)*. Springer, 614–621.

S. Tausch, S. Ta, and H. Hußmann. 2016. A comparison of cooperative and competitive visualizations for co-located collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 5034–5039.

J. Terken and J. Sturm. 2010. Multimodal Support for Social Dynamics in Co-located Meetings. *Personal and Ubiquitous Computing* 14, 8, 703–714.

P. Thagard. 1989. Explanatory coherence. *Behavioral and brain sciences* 12, 03, 435–467.

E. L. Thorndike. 1913. *Educational psychology*. Vol. 2. Teachers college, Columbia University.

K. Tollmar, O. Sandor, and A. Schömer. 1996. Supporting social awareness @ work design and experience. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW '96)*. ACM, 298–307.

E. P. Torrance. 1974. *Torrance Tests of Creative Thinking: Normstechnical Manual: Research Ed.: Verbal Tests, Forms A and B: Figural Tests, Forms A and B. Flere Materialer*. Personell Press.

E. P. Torrance and K. Goff. 1989. A quiet revolution. *The Journal of Creative Behavior* 23, 2, 136–145.

S. E. Toulmin. 1958. *The uses of argument*. Cambridge University Press.

P. Tunstall and C. Gsipps. 1996. Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal* 22, 4, 389–404.

J. S. Valacich, A. R. Dennis, and J. F. Nunamaker. 1992. Group-Size and Anonymity Effects on Computer-Mediated Idea Generation. *Small Group Research* 23, 1, 49–73.

J. S. Valacich, B. C. Wheeler, B. E. Mennecke, and R. Wachter. 1995. The effects of numerical and logical group size on computer-mediated idea generation. *Organizational Behavior and Human Decision Processes* 62, 3, 318–329.

D. Van-Dijk and A. Kluger. 2001. Goal orientation versus self-regulation: Different labels or different constructs. (2001). Paper presented at the 16th annual convention of the Society for Industrial and Organizational Psychology, San Diego, CA.

D. Van-Dijk and A. N. Kluger. 2004. Feedback sign effect on motivation: Is it moderated by regulatory focus? *Applied Psychology* 53, 1, 113–135.

F. H. Van Eemeren and R. Grootendorst. 1992. *Argumentation, communication, and fallacies: A pragma-dialectical perspective.* Lawrence Erlbaum Associates, Inc.

F. H. Van Eemeren, R. Grootendorst, R. H. Johnson, C. Plantin, and C. A. Willard. 1996. *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Routledge.

A. B. VanGundy. 1988. *Techniques of structured problem solving*. Van Nostrand Reinhold New York.

K. Verbert, S. Govaerts, E. Duval, J. L. Santos, F. Van Assche, G. Parra, and J. Klerkx. 2014. Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing* 18, 6, 1499–1514.

F. B. Viégas and J. S. Donath. 1999. Chat circles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, 9–16.

C. von Aufschnaiter, S. Erduran, J. Osborne, and S. Simon. 2008. Arguing to learn and learning to argue: Case studies of how students' argumentation relates to their scientific knowledge. *Journal of Research in Science Teaching* 45, 1, 101–131.

L. S. Vygotsky. 1980. *Mind in society: The development of higher psychological processes*. Harvard University Press.

D. Walton and E. C. Krabbe. 1995. *Commitment in dialogue: Basic concepts of interpersonal reasoning*. SUNY Press.

H.-C. Wang, D. Cosley, and S. R. Fussell. 2010. Idea expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. ACM, 103–106.

H.-C. Wang, S. R. Fussell, and D. Cosley. 2011. From diversity to creativity: Stimulating group brainstorming with cultural differences and conversationally-retrieved pictures. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. ACM, 265–274.

A. Weinberger, K. Stegmann, and F. Fischer. 2010. Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior* 26, 4, 506 – 515.

A. Weinberger, K. Stegmann, F. Fischer, and H. Mandl. 2007. Scripting argumentative knowledge construction in computer-supported learning environments. In *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives*. Springer, 191–211.

B. Weiner. 1974. *Achievement motivation and attribution theory*. General Learning Press.

S. P. Weisband, S. K. Schneider, and T. Connolly. 1995. Computer-mediated communication and social information: Status salience and status differences. *The Academy of Management Journal* 38, 4, 1124–1151.

A. Wessels, S. Fries, H. Horz, N. Scheele, and W. Effelsberg. 2007. Interactive lectures: Effective teaching and learning in lectures using wireless networks. *Computers in Human Behavior* 23, 5, 2524–2537.

N. Wiener. 1948. *Cybernetics: or control and communication in the animal and the machine.* The MIT Press.

M. Wilkerson, W. G. Griswold, and B. Simon. 2005. Ubiquitous presenter: Increasing student access and control in a digital lecturing environment. In *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education (SIGCSE '05)*. ACM, 116–120.

F. Winter. 2011. *Erweiterung und Evaluation eines Frameworks für Feedback-Systeme zur Unterstützung von Team-Sitzungen in der gymnasialen Oberstufe*. Diploma thesis.

R. Xiong and J. Donath. 1999. Peoplegarden: Creating data portraits for users. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology (UIST '99)*. ACM, 37–44.

N. Yoshida, S. Fukushima, D. Aida, and T. Naemura. 2016. Practical study of positive-feedback button for brainstorming with interjection sound effects. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, 1322–1328.

S. A. Yourstone, H. S. Kraye, and G. Albaum. 2008. Classroom questioning with immediate electronic response: Do clickers improve learning? *Decision Sciences Journal of Innovative Education* 6, 1, 75–88.

R. B. Zajonc. 1965. *Social facilitation*. Research Center for Group Dynamics, Institute for Social Research, University of Michigan.

J. Zhou. 1998. Feedback valence, feedback style, task autonomy, and achievement orientation: Interactive effects on creative performance. *Journal of Applied Psychology* 83, 2, 261.

J. Zhou. 2003. When the presence of creative coworkers is related to creativity: Role of supervisor close monitoring, developmental feedback, and creative personality. *Journal of Applied Psychology* 88, 3, 413 – 422.

B. Zhu. 2002. *Information visualization for knowledge repositories: applications and impacts*. Ph.D. Dissertation.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

München, den 09. August 2016

Sarah Tausch