Metaphor or Diagram? Comparing Different Representations for Group Mirrors

Sara Streng\textsuperscript{1}, Karsten Stegmann\textsuperscript{2}, Heinrich Hußmann\textsuperscript{1}, and Frank Fischer\textsuperscript{2}
\textsuperscript{1}Ludwig-Maximilians-Universität \\
Department of Computer Science \\
Amalienstr. 17, 80333 München \\
\{sara.streng, heinrich.hußmann\}@ifi.lmu.de

\textsuperscript{2}Ludwig-Maximilians-Universität \\
Department of Psychology and Education \\
Leopoldstr. 13, 80802 München \\
\{karsten.stegmann, frank.fischer\}@psy.lmu.de

\textbf{ABSTRACT}
This paper aims at answering the question how ambient displays can be used as group mirrors to support collaborative (learning) activities. Our research question is to what extent the type of feedback representation affects collaborative processes. Two different representations have been created and compared in a user study: a diagram and a metaphor. In the diagram version the quality rating for each person is explicitly shown in charts and numbers. In the metaphorical representation feedback is implicitly visualized by changing certain characteristics of a pictorial scene. The results show that the metaphoric group mirror was not only more popular than the diagram, it also had a greater impact on the group behavior. When receiving negative feedback from the metaphoric group mirror, a correction of behavior was made significantly faster than with the diagram. Furthermore, both group mirrors had a positive effect on the self-regulation of the group compared to the baseline condition without feedback.

\textbf{Author Keywords}
Group Mirror, ambient display, metaphor, collaborative learning.

\textbf{ACM Classification Keywords}
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

\textbf{INTRODUCTION}
Group mirrors (or “mirroring systems”) provide feedback to groups by reflecting certain aspects of collaborative activities (Jermann et al., 2001, p. 3), e.g. the amount of participation of individuals to a group discussion. They differ from awareness tools in that the collaborating group members are physically co-located and actively interacting at that time. On the contrary, awareness tools typically show information about co-workers who are not present, such as whether they are available or which activities they are currently engaged in. Examples for such awareness tools are Portholes (Dourish and Bly, 1992), NYNEX Portholes (Lee et al., 1997) or ActiveMap (McCarthy and Meidel, 1999).

Feedback plays an important role in collaborative processes. Usually collaborators self-regulate their behavior according to feedback they get during collaboration. In an argument, for example, a participant will add additional backing for a claim if another participant shakes her or his head. Furthermore, several empirical studies showed a positive effect of feedback on knowledge acquisition in learning scenarios (for a review see Hattie and Timperley, 2007).

Group mirrors can be used to provide feedback in both, collaborative work settings (e.g. DiMico et al., 2004) and collaborative learning scenarios (Jermann et al., 2001). One reason why they have been applied by educationalists is that they believe the learning outcome can be maximized by balancing participation levels (Bachour et al., 2008). Although a moderator or teacher can also regulate things like participation shares, interruptions by a third party are often too obtrusive. Group mirrors on the other hand may provide feedback in a subtler and less interruptive way. Thereby, feedback by group mirrors may offer the chance to self-regulate activities with a minimum of disturbance of the collaboration itself. However, if the feedback provided by the group mirror requires too much attention, collaborators may be distracted from the task as well as require more time to react to the feedback.

Previous work has shown that group mirrors can affect group behavior (see Related Work). However, group mirrors have many different characteristics, of which not all have been equally examined so far. In this work, the effect of different feedback representations is examined. An ambient display is used to mirror the quality of argumentation in small groups. Two different

\textbf{Figure 1. Metaphoric Group Mirror in Use.}
representations are compared with a baseline condition without group mirror in a controlled experiment. First, a diagram, which is commonly used for group mirrors, explicitly shows feedback in numbers. Second, a metaphoric representation visualizes the same data by changing parts of a pictorial scenery as shown in Figure 1. The feedback is implicitly shown by the weather, which can range from cloudless sky to heavy rain, as well as the condition of the trees, which blossom when the quality of argumentation is very good or lose their leaves when the performance is poor. In the baseline condition no group mirror is used at all.

The research question is to what extent the type of representation (diagram vs. metaphor vs. no feedback) affects the quality of collaboration. Our expectation is that both feedback representations support the self-regulation of collaborative processes in comparison to the baseline group without group mirror as a matter of feedback. Furthermore, we expect the metaphoric representation to better support the self-regulation of collaborative processes in comparison to the diagram version due to the lower attention required to grasp the feedback from the metaphoric representation.

The paper is structured as follows: At first, related work on group mirrors is summarized and examined for commonalities, differences and limitations. Based on this, different characteristics of group mirrors are established, which help to classify different types of mirroring systems, such as the type of visualization, the placement of the mirror and the type of information. The latter one can be further divided into qualitative and quantitative feedback, which are subsequently discussed in more detail. Afterwards so-called collaboration scripts are introduced that define ideal collaborative activities in terms of tasks, roles and phases. In the empirical study, the feedback provided by the group mirror referred to the compliance to two of these scripts. Then, the different feedback representations are described. At this time the quality of the group process has to be entered manually using a control (or Wizard of Oz) interface. This interface is illustrated in the succeeding section. Subsequently, the method of the user study is described and the results are summarized and interpreted. Finally, we discuss our results and draw conclusions on the application of our findings to develop effective group mirrors.

RELATED WORK
DiMicco et al. (2004) studied the effect of showing speaker rates on a shared wall display during a decision-making process. Their findings were that the presence of this group mirror made over-participants significantly decrease the amount they spoke. Under-participants, however, did not respond by increasing their participation. A table-based system with a similar purpose is Reflect (Bachour et al., 2008), a noise-sensitive table, which reflects the current participation levels in the form of colored LEDs. Both visualizations are highly comparative. The Conversation Clock (Bergstrom and Karahalios, 2007) shows the same information in a less comparative way. Current as well as previous participation levels (measured by microphones) are indicated by colored bars, which are arranged in concentric rings. This visualization focuses rather on the history than providing the means for a comparison between the individuals. Furthermore, there are systems (e.g. Mathur and Karahalios, 2009) that visualize remote conversation; however, the idea is rather providing graphical log files for future use than influencing group behavior.

Many group mirrors focus on the easy to measure but also rather superficial level of participation shares. There are some exceptions, e.g. the Conversation Votes table, which displays anonymous positive or negative votes in addition to the participation levels. We identified feedback on the quality of collaborative activities by group mirrors as an important but not well-examined field of research. Another aspect of the development of group mirrors that has not been tackled systematically by researchers so far is the question how the feedback is represented. Existing group mirrors often use diagrams or advanced visualizations (such as concentric rings). Metaphors are sometimes used in ambient display applications (e.g., Nesbitt and Shen, 2007, Redström et al., 2000 or Ishii et al., 1998), but have not been used for group mirrors.

The different characteristics of group mirrors, which were implicitly mentioned above, are analyzed on an abstract level in the next section.

CHARACTERISTICS OF GROUP MIRRORS
Different types of group mirrors can be classified according to the following criteria: (1) visualization (how?), (2) placement (where?), and (3) type of mirrored information (what?).

Type of Visualization
There are various types of visualizations and several taxonomies, which classify them according to different criteria (e.g., Lohse et al., 1994, Shneiderman, 1996 or Lengler and Eppler, 2007). Shneiderman identified seven data types (1-, 2- and 3-dimensional, temporal, multi-dimensional, tree and network) and interface design issues that come with each of them. The information that is represented by group mirrors needs to be very simple or otherwise too much attention may be needed to grasp the feedback. Hence, the group mirror would be too distracting and time-consuming to interpret. Consequently, they are confined to 1- or 2-dimensional data. For the same reason the amount of data is fairly low (e.g. only one participation level per person). The most straightforward representation for such data is a diagram. That is why most existing group mirrors use some form of diagram, be it as a rendered chart or a matrix of light bulbs or LEDs. An alternative for visualizing such simple and little information is using metaphors. For example, the information could be translated into colors of an ambient display or manipulate certain aspects of a pictorial scene.

Placement
Group mirrors are typically either placed on walls or integrated in a table. Both have pros and cons. Tabletop displays have the advantage that they are physically in the
center of the group, so that everybody is facing the display. Regarding the content there is, however, an orientation problem because people are usually sitting around the table. Consequently for some people the view is upside-down. Wall displays do not have this problem, but in a typical seating arrangement for small groups, the display is not in everybody’s visual field.

**Type of Information**

Finally, the type of information can be further divided into quantitative and qualitative information. Quantitative data is by far easier to capture automatically through sensors. This is the main reason why most systems nowadays measure participation rates and other quantitative information. An example for qualitative feedback is mirroring the importance of a person’s arguments, instead of the number or even just the time that was spoken. The next section will discuss this issue in more detail.

**MIRRORING THE QUALITY OF COLLABORATIVE PROCESSES**

As pointed out by Dillenbourg (2005, p. 255) the usefulness of mirroring pure quantitative data is questionable. Using the example of participation rates, while unbalance participation is an indicator for a deficient (learning) outcome, the opposite is not necessarily true. In order to leave a good impression on the group mirror, participants could even produce useless arguments, which would undermine the purpose. Therefore, group mirrors should ideally (also) offer feedback on the quality of collaborative processes. However, qualitative data is by far more difficult to acquire. While directing a microphone to each person allows for an easy capturing of participation shares, it is much more complicated to assess the quality of collaborative processes – even by humans. How could a computer cope with that problem?

In the following at first quality criteria for evaluating group processes have to be defined. In a second step current possibilities of automatic analyses of collaborative processes as well as scenarios without automatic analysis are discussed.

**Defining Quality Criteria: Using Collaboration Scripts as Criterion for Quality of Collaborative Processes**

To be helpful feedback has to include information on difference between the target state and current state. The target state has to be known by the feedback receiver. The question is, what can be regarded as target state with regard to collaborative processes? In the context of the learning science, so-called collaboration scripts (e.g. Stegmann et al., 2008) can be used to define ideal collaborative processes.

Collaboration scripts are instructional plans that specify and sequence collaborative learning activities. When needed, these scripts assign various activities to the different learners (Kobbe et al., 2007). Collaboration scripts typically focus on those activities that are regarded as facilitating knowledge acquisition (King, 2007). The script approach is not limited to specific activities, but may focus on those activities regarded as beneficial for learning (cf. Kobbe et al., 2007; Kollar et al., 2007)

and/or aim to reduce extraneous activities like off-topic talk (e.g. Baker and Lund, 1997). Hence, each collaboration script is explicitly or implicitly based on an approach to collaborative learning that specifies crucial cognitive and social processes as well as associated activities, learning goals, and the relationship between these elements.

The collaboration script that was used to structure the experiment session in this case was the so-called peer-review script (e.g. Weinberger, 2003). It distributes the roles of problem analysts and critics and specifies the activities for each role. The problem analyst has to provide analyses of the problem and has to make proposals how to solve the problem. The critics have to provide constructive critique including alternative proposals. Each activity has a defined duration (phases) and should not be disturbed by other participants (except for clarification questions). Usually, the script also includes a rotation of roles, but for reasons of simplicity role reversal was abstained from in this experiment.

Another script that played a role in the experiment setting was the script for the construction of single arguments (Stegmann et al., 2008). It specifies how to construct arguments during a discussion. The script asks participants of a discussion to provide grounds and qualifications for their claims. The claim expresses the position on an argument. The elements that support the claim regarding Toulmin’s (1958) model, which we subsume under the term grounds, are data, warrant and backing. Data involves factual information such as an observation, which supports the acceptance of the claim. A warrant justifies the inference between data and claim. Warrants are usually theoretical laws, rules, or definitions. Backing is evidence such as statistics or expert opinions that is in line with the warrant. Qualifiers and their associated rebuttal are the elements in Toulmin’s (1958) model that limit the validity of a claim. We use the term qualifications for these elements, because they qualify the relationship between claim and warrant. Toulmin’s (1958) qualifier reflects uncertainty with regard to the validity of the claim and is usually expressed using modal adverbs such as “perhaps” or “probably”. While the qualifier only expresses a potential limitation, the so-called rebuttal describes the state of affairs when the claim is invalid.

**(Automated) Assessment of Quality of Collaborative Processes**

Current interdisciplinary approaches of learning scientists and computer linguists show that the automated assessment of quality of collaborative processes is already possible with high accuracy for written discussions. Rosé and colleagues (2008), for example, trained automated classifiers for different dimensions of qualities of collaborative processes, e.g. content or quality of argumentation. The authors used classified written online discussions and trained automated classifiers with this data. Therefore, a system that transcribes spoken discussions and applies such automated classifiers to the transcribed discussion does not seem to be unrealistic.
However, it is still a long way to a reliable qualitative real-time analysis of spoken arguments.

In meantime, a moderator or teacher can also do the assessment by manually rating the quality of arguments. The advantage of using a group mirror instead of spoken feedback is the unobtrusiveness. The tutor can silently give feedback without having to interrupt the ongoing discussion.

IMPLEMENTATION OF THE GROUP MIRROR
The prototypical implementation that was used in the experiment consisted of two connected notebooks. One of them showed the group mirror representation, which was projected to a wall. The other notebook was used by the wizard to manually enter the quality assessment that was visualized by the group mirror. Both graphical user interfaces were implemented using Flash and are presented in detail in this section.

Group Mirror Representations
The group mirror representations has to match the quality criteria defined in the collaboration script the participants act on. In this respect we agree with other literature that came to the same conclusion regarding awareness tools (Espinosa et al., 2000), which is that they highly depend on the task at hand. Therefore, we first describe the collaboration script that participants were asked to follow before presenting the implementation of the two different representation types.

The two other participants had the role of critics. In one minute each, the critics should provide their major concerns regarding the position of the analyst. Afterwards, the analyst had again two minutes to answer the critics and to revise her or his statement. The critics again should provide a constructive review (one minute each), before the analyst had to provide her or his final position statement. All participants are advised not to speak (except for clarification questions) if it is not their turn. Furthermore, the participants were, regardless of their role within the peer review script and according to the script for the construction of single arguments, asked not to provide bare claims, but to provide reasons and qualifications for their claims.

Accordingly, the group mirror has to provide feedback on the compliance with the peer review script (turn taking, untimely contributions, duration of contribution) and the quality of argumentation (bare claim vs. grounded claims with qualification). In the following two different visualizations are presented to provide this feedback. One is a diagram, which is commonly used for group mirrors. The other is a metaphorical representation of the same information.

Metaphoric Group Mirror
Figure 2 shows five screenshots of the metaphoric group mirror. It was designed to provide feedback to three people regarding the quality of their argumentation. The analyst’s quality of argumentation is represented by the weather. There are five stages from cloudless sky (the analyst is providing grounded claims with qualifications, as in picture 1) to strong rain (the analyst is providing bare claims, as in picture 3). The analyst’s phase starts with sunrise and ends with sunset. However, the timeline is just a suggestion. If the analyst takes less time, the sunset is accelerated. If the analyst takes more time, the sunset “waits” until she or he finished her or his turn.

Each of the two critics is represented by one tree. The names of the critics are placed below “their” tree. As for the analyst there are also five stages to provide feedback on the quality of argumentation. If a critic provides grounded claims with qualifications the corresponding tree is flourishing (as in picture 4), if the critic provides just bare claims, the tree loses its leaves (see picture 5).

The critics share the night-time, which means that the handover should happen when the moon is at the uppermost point. Again, this is only a lead, not an instruction. Initially average ratings are shown: partly cloudy sky and trees with leaves but without blossoms (both matching a 3 out 5 stages as in picture 2).

A sudden lightning covers the whole display when one participant interrupts another (see Figure 3). The lightning lasts until the interruption ends. As the lightning covers the whole display, the emitted warning is directed towards the whole group whereas all other information is mapped to a specific part (tree or weather) and thus to a specific person. However, we expect that the interrupter is likely to feel addressed no matter whether the bolt will hit one of the trees or not.
In order to be able to compare the effectiveness of diagram and metaphor, the mirrored information must be the same for the diagram. The quality of people’s arguments are shown in a column chart (see Figure 4). Each of the three columns represents one person. The five rating stages are centered around the neutral value 3. Below the bar is a count down that shows the time intended for the corresponding participant. Of course, only one of three count downs is active at a time.

Control Interface
The group mirror can be controlled with a simple user interface as shown in Figure 5. There are two different use cases for the control interface. In learning settings, a teacher or tutor can use it to silently give feedback to a group of learners. In this case, the learners know that the ratings come from their tutor. For the purpose of this experiment, however, the Wizard of Oz approach (Kelley, 1983) was used. In Wizard of Oz experiments participants think they are interacting with a fully implemented system, while in reality there is a human being, who manually takes over part of the functionality. Correspondingly, participants of this experiment thought that the experimenter takes notes on the laptop. They were not told that the experimenter is the Wizard who rates the quality of their argumentation and their compliance to the peer review script.

In the following control (or Wizard of Oz) interface is described. In the upper part the quality of argumentation for each of the three participants can be entered on a scale from 1 (grounded and qualified claims) to 5 (bare claims). The ratings are automatically send to the group mirror every 10 seconds. If the wizard/tutor wants the mirror to be updated earlier than that, the “send now” button in the bottom right can be used. After pressing it, the new ratings are transmitted immediately and the timer is reset to 10 seconds. There are two more buttons in the center. The “untimely contribution” button turns the lightning or blinking red circle on and off, which signals an interruption by one of the participants. The “rush” button can be used to accelerate sun- or moonset or respectively the countdown when the current speaker takes less time than scheduled by the mirror.

EMPIRICAL STUDY
Research Questions
A study was conducted to evaluate the effectiveness of the group mirror in general as well as the different visualizations. The research questions of our empirical study are:

(1) To what extent does the type of representation (without vs. diagram vs. metaphor) facilitate the self-regulation of participants of a discussion?

(2) To what extent does the type of representation (without vs. diagram vs. metaphor) affect the subjective estimation of compliance to the collaboration script?

(3) Which type of representation (diagram vs. metaphor) is preferred by the participants?

Method
Each discussion was held with groups of three participants. Altogether 27 people at the age of 19 to 53 participated in the study. A one-factorial design was used with three conditions. We varied the factor “type of representation” within subjects (without vs. diagram vs. metaphor), i.e. the same group of participants discussed under all three conditions. While the “without representation” (baseline) was always first, the order of diagram and metaphor was randomized. The roles of the peer review script, however, were fixed. In the control condition no mirroring system was used at all. During the condition with group mirror, the diagram or the metaphor are projected onto a wall next to table where the group of
three discussed. No participant sat with the back to the representation.

The topics that were chosen for discussion have attracted public attention in Germany over a long period of time. Hence, no additional material (and preparation time) was needed to discuss these topics. The topics were (1) data security and online search, (2) nuclear and green energy and (3) speed limits and carbon emission.

After welcome, the experimenter shortly explained the general scope of the study without mentioning the concrete research questions and hypotheses. The experimenter further gave a short overview on the different steps of the study. Before the participants were asked to discuss the first topic, they were given a short introduction to the collaboration script and the quality criteria. In addition, before each condition with group mirror the experimenter handed out a short presentation on how the different process qualities are represented in the group mirror. After each discussion, the participants were asked to fill out specific questionnaires regarding the previous condition.

We measured self-regulation of collaborative processes by computing the time between a negative feedback and the moment when the process was improved (or the end of a turn was reached). In the baseline condition without group mirror, the quality of the process was assessed in the very same way, but there was no feedback displayed. After each discussion with group mirror a questionnaire was filled out that asked for the subjective estimation of compliance to the role as well as compliance to the timing. Finally, after all three conditions the students were asked which of the two types of representation they preferred.

The three conditions were compared using the General Linear Model (GLM) for repeated measures. To compare contrasts post-hoc between conditions without vs. with group mirror or between the diagram condition vs. the metaphor condition, t-Tests for dependent variables were applied. All statistical tests are applied on a 5% level of significance. Effect sizes are either reported as $\eta^2$ (GLM) or $d$ (t-Test).

**Results**

(RQ1) To what extent does the type of representation (without vs. diagram vs. metaphor) facilitate the self-regulation of participants of a discussion?

The statistical analyses show that the type of representation had a significant and strong effect on the self-regulation of collaborative processes, $F_{(2)} = 8.63; p < .05; \eta^2 = .25$. Post-hoc comparisons show that the self-regulation was significantly slower in the baseline condition compared to both, the diagram condition, $t(26) = 2.74; p < .05$; two-tailed; $d = .53$, as well as the metaphor condition, $t(26) = 3.16; p < .05$; two-tailed; $d = .62$. Furthermore, the self-regulation was significantly slower in the diagram condition compared to the metaphor condition, $t(26) = 1.98; p < .05$; two-tailed; $d = .39$ (see Figure 6).

(RQ2) To what extent does the type of representation (without vs. diagram vs. metaphor) affect the subjective estimation of compliance to the collaboration script?

Regarding subjective estimation of compliance to the role, the GLM shows a significant effect with a medium effect size, $F_{(2)} = 3.10; p = .05; \eta^2 = .11$ (see Table 1). Post-hoc comparisons show that the subjective estimation of compliance to the role was significantly better in the diagram condition compared to the baseline condition, $t(26) = 2.06; p = .05$; two-tailed; $d = .39$. Neither the difference between the baseline condition and the metaphor condition, $t(26) = 1.77; n.s.;$ two-tailed, nor the difference between the diagram condition and the metaphor condition, $t(26) = -.94; n.s.;$ two-tailed, were significant.

(RQ3) Which type of representation (diagram vs. metaphor) is preferred by the participants?

Finally, the participants were asked which visualization they would choose when given the choice. A significant majority of 70%, $\chi^2(3) = 30.93; p < .05$; two-tailed,

![Figure 6. Time taken from a negative feedback until a measured improvement (in seconds).](image)
would prefer the metaphor (see Figure 7). The most frequent reasons given were that it was “more vivid”, “fun”, “aesthetic”, “easier to understand”, “friendlier” and “better noticeable”. Another positive comment concerned the lack of a clear distinction between positive and negative ratings, which was regarded as “pedagogically valuable”.

**Figure 7. Preference for Visualization Types.**

**DISCUSSION**

One motivation for doing this research was finding out more about the perception of information on ambient displays. According to Mankoff et al. “ambient displays are abstract and aesthetic peripheral displays portraying non-critical information on the periphery of a user’s attention” (Mankoff et al., 2003, p. 1). Although numerous ambient display applications have been developed, there has been little in-depth evaluation on them. A probable explanation is that they are designed to be unobtrusive, which makes it difficult to measure an effect. Often, the interaction with ambient displays occurs spontaneously and too rare to allow for an in-depth evaluation. By applying collaboration scripts as demonstrated in this paper, the interaction can be condensed to a session of 10 minutes. Thereby it is not only easier to observe the effects of the group mirror - it is also possible to compare different visualizations in a controlled experiment.

However, there are two drawbacks. First, as in every lab experiment, there is the danger that the artificial situation distorts the results. Regarding the comparison of diagram and metaphor there is no reason to expect an imbalanced bias. In other words, if there is a distortion, it will most likely affect both visualizations the same way. The second more severe problem is the question of generalizability. In order to be able to conduct the experiment, a very specific situation was chosen in terms of tasks, roles, phases and group size. While the metaphoric visualization has clear advantages in this case, the question is whether the results can be generalized to all group mirrors or even all awareness tools or ambient displays? The answer is most likely that the preference of a metaphoric visualization largely depends on the type of information and the quality of visualization.

Nevertheless, this paper demonstrates the maybe underestimated potential of metaphoric visualizations: They can communicate feedback in a subtler and more playful way while still catching the eye by encoding events or changes in the assessment in order that larger parts of the screen are affected.

**CONCLUSIONS AND FUTURE WORK**

In this paper an ambient display application was presented, which reflects the quality of argumentation in small groups. This so-called group mirror was implemented in two versions. The first version visualized feedback in the form of diagrams (using bar charts, numbers and a legend). The second version metaphorically encoded the same information in a scenery. Thereby, the quality of the individual participants’ arguments influenced the weather and the condition of two trees in the picture.

Two collaboration scripts have been presented, which played a role in the experiment setting. Collaboration scripts are educational instruction plans that specify and structure collaborative learning activities. The first script is the peer-review script, which distributes the roles of a problem analyst and critics among the learners. In this experiment, there was one analyst and two critics, who took alternating turns. The analyst starts by analyzing a problem. Subsequently, the two critics provide constructive feedback, which the analyst can then consider in his revision, and so forth. Secondly, the script for the construction of single arguments was used, which asks the participants to provide grounds and qualifications for their claims.

An empirical study was conducted using a one-factorial design with three conditions. In each condition a different type of feedback representation was chosen: (1) a diagram, (2) a metaphor and (3) no feedback. The results show that both types of feedback representations had a positive effect on the self-regulation of group processes as compared to the baseline condition without feedback representation. Furthermore, the metaphorical version of the feedback facilitated the self-regulation of group processes significantly better than the diagram version. Deficient behavior was corrected significantly faster when receiving negative feedback from the metaphoric group mirror as compared to the diagram condition. Furthermore, a majority of 70% would choose the metaphoric group mirror over the diagram.

These results demonstrate the potential of metaphoric visualizations. Although the transferability of the results to other scenarios largely depends on many aspects (e.g. the type of information and the quality of the visualization), it can be assumed that the advantages of metaphors apply to a wider range of settings than the specific scenario that was used here. To confirm this, we plan to conduct more studies using different collaborative tasks and representations. Furthermore, a systematic comparison of representations is necessary to link specific characteristics (such as the size of the display area used to send a signal) to the effectiveness.

**ACKNOWLEDGMENTS**

We would like to thank Younes Alj for his help with designing the group mirror representations and conducting the experiment.
REFERENCES


