Running head: SUPPORTING REMOTE PROBLEM-SOLVING

Research Article

Supporting Remote Collaborative Problem-Solving

Miriam Hansen¹ and Hans Spada²

¹Department of Psychology, Goethe-University Frankfurt

²Department of Psychology, University of Freiburg
ABSTRACT

We propose a framework of individual problem-solving and communicative demands (IproCo) that bridges the gap between models from cognitive psychology and communication pragmatics. Furthermore, we present two experiments conducted to identify factors influencing the demands and to test possibilities for support. The experiments employed a remote collaborative picture-sorting task with concrete and abstract pictures and applied non-interactive conditions compared to interactive conditions. In a first experiment, the influence of the postulated demands on collaboration process and outcome was analyzed, and the impact of shared applications was tested. In a second experiment, we evaluated instructional support measures consisting of a model collaboration and a collaboration script. The collaboration process showed benefits of the support but the outcome did not. However, the support measures fostered the collaboration process even in the particularly difficult conditions with non-interactive communication. We discuss the impact of the IproCo framework and apply it to other tasks.

Keywords: remote collaboration, referential communication, change blindness, shared application, collaboration script
Supporting Remote Collaborative Problem-Solving

Picture the following situation: A production worker realizes the engine he is working with is no longer running. The mechanic who is normally in charge of repairing the engine is not available. The worker tries to identify the problem, but as the engine is large and highly complex, consisting of a huge amount of individual parts, he is unable to find it. The worker knows about a second engine of this type in another enterprise location and makes a phone call to a mechanic there. With the help of the mechanic’s descriptions, he hopes to find the broken part and repair the engine. Will his plan be successful?

This example represents a broad range of situations requiring remote collaboration. With the development of increasingly sophisticated technologies for remote collaboration (e.g. videoconferencing technology, Internet-based meeting technology), the occurrence of such situations has increased substantially in different working contexts, as has the amount of research conducted in the field of remote collaboration (Olson & Olson, 2000; see also Fussell, Kraut & Siegel, 2000; Rummel & Spada, 2005).

Collaboration in remote settings is very challenging. In the situation described above, the worker’s success will depend on how well the worker and the mechanic deal with the demands placed on them. They have to face individual cognitive demands such as visual processing of the engine, including visual perception, object identification, and visual search (we will refer to these demands as *individual problem-solving demands*). Furthermore, the collaborators also face *communicative demands*: They need to find a way to describe to each other what their engines look like in order to compare them and to find the problem in the broken one. To assure mutual understanding of the expressions used to refer to the engine parts, the collaborative process of grounding has to take place (Clark & Brennan, 1991; Clark & Schaefer, 1987). Grounding is
notably difficult in remote settings due to the reduced or lack of physical copresence (Clark & Brennan, 1991; Kraut, Fussell & Siegel, 2003). The worker and mechanic are only interacting via an audio channel, and are neither able to see each other’s facial expression nor use deictic gestures for grounding (Clark & Krych, 2004; Fussell et al., 2004). Due to the various demands on the collaborators, errors and problems are likely to occur and a satisfactory solution may not be found within a short period of time. To fulfill the worker’s desire for a good collaboration outcome, their collaboration should be supported.

Such support may be realized by improving the technical environment (e.g. adding a visual channel; Whittaker & O’Connaill, 1997; providing shared workspaces; Tang, 1991; Whittaker, Geelhoed & Robinson, 1993) or by structuring the interaction process (e.g. collaboration scripts; O’Donnell & Dansereau, 1992; Baker & Lund, 1997). Collaboration scripts have also been used to instruct collaboration partners on how to solve a remote collaborative task (Rummel & Spada, 2005). With this instructional approach, a sustainable enhancement of the collaborative process and results can be achieved. However, in order to design such instructional support measures, we need to understand in detail what skills are needed to solve such a collaborative task.

AIMS OF THE PAPER

With this paper, we aim to give a detailed picture of the relevant processes of collaboration in our new remote picture-sorting task. We try to identify the individual problem-solving and communicative demands of the task in order to develop effective instructional support measures. Most research on collaboration and communication focusing on real-world contexts does not take the findings of cognitive science sufficiently into account (see McNamara, 2006). A promising exception is the model recently presented by Horton and Gerrig (2005a),
which uses cognitive theories of memory to explain grounding and audience design processes. Several studies have illustrated how this memory-based approach helps to improve the understanding of communication processes (Horton, 2007; Horton & Gerrig, 2005b; Horton & Spieler, 2007).

To build a foundation for instructional support, we also try to bridge the gap and bring cognitive models together with models of communication. Therefore, the first part of this paper presents a theoretical framework of the task demands, emphasizing the individual problem-solving demands and the links to research on communication. In a first experiment, we examined the impact of these theoretically derived demands on collaboration process and outcome.

Our second aim is to evaluate support measures for the remote collaborative picture-sorting task. In the first experiment, we tested simple technical support measures (shared applications), while in the second experiment, we evaluated instructional support measures (model collaboration and collaboration script) including support of the individual problem-solving and the communicative demands.

THE COLLABORATIVE PICTURE SORTING TASK

Studies of collaboration mostly use realistic and therefore quite complex settings (e.g. Fussell et al., 2000; Olson & Olson, 2000; Rummel & Spada, 2005) and do not allow the underlying processes to be examined in depth. Therefore, we used a more restricted remote collaborative task that enables a more detailed analysis. In the following, we briefly describe the task before presenting the framework of demands.

In the broadest sense, the task used in our studies can be classified into a “general class of ‘mentoring’ collaborative physical tasks” (Kraut et al., 2003; p. 16), in which one person manipulates pictures and receives instructions from another person about where to place them.
The task accomplished by the worker and mechanic in the example described above also fits into this category. The structure of these tasks is similar to the referential communication task (Krauss & Weinheimer, 1966), which has often been used to study communication (e.g. Brennan & Clark, 1996; Horton & Keysar, 1996). In a referential communication task, a speaker typically describes some clearly different pictures to an addressee. In previous referential communication studies, impacts of task characteristics as codability, discriminability, and familiarity of the material (Bortfeld, Leon, Bloom, Schober & Brennan, 2001; Gergle, Kraut & Fussell, 2006; Horton & Gerrig, 2002; Hupet, Seron & Chantraine, 1991) have been found to have an influence on the communicative effort needed to achieve mutual understanding.

However, our collaborative picture-sorting task differed on one important point from the classical task: Our participants needed to conduct a visual search to detect relevant differences between quite similar pictures. Thereby, individual problem-solving demands (visual search processes) were added to the communicative demands: As in realistic collaborative physical tasks, this corresponded to a dual task situation in which the participants had to accomplish individual problem-solving activities in parallel to communication. In contrast to previous work, our task allowed to scrutinize the impact of task characteristics on both individual problem-solving and communicative demands.

In both experiments included in this paper, two persons sitting in two different rooms jointly solved a picture-sorting task. One of the participants assumed the role of speaker and the other took the role of addressee (synonymous terms are director and matcher). In the following, we adopt the convention of assuming a female speaker and a male addressee. The task was presented on two displays and oral communication between speaker and addressee was possible via an audio link.
On the speaker’s display, a number of pictures were presented that differed only in terms of minor details. The speaker had to detect the relevant features and describe them and their order to the addressee. The addressee saw the pictures in a random order and had to arrange them according to the speaker’s description. The addressee could rearrange the pictures on the target area by using the mouse (drag and drop). Because the differences between the pictures were very small, the participants first had to search for them. This component of feature detection accomplished in parallel to communication constitutes the main difference from the classical referential communication task, in which the task demand consists in the verbal description of clearly different pictures or objects. This additional individual problem-solving demand makes the task more comparable to realistic collaborative tasks in which communication often has to take place in parallel to individual problem-solving processes (e.g. the worker communicating on the phone with the mechanic while searching for a broken part, or an air traffic controller monitoring different displays while interacting with pilots). Furthermore, the task is comparable to collaborative settings with distributed resources or distributed skills, such as expert-layperson communication (e.g. Bromme, Jucks, & Wagner, 2005; Nückles & Ertelt, 2006) or hidden-profile tasks (e.g. Stasser & Titus, 1985), because the speaker has information about the required target order of the pictures that the addressee does not have.

FRAMEWORK OF INDIVIDUAL PROBLEM-SOLVING DEMANDS AND COMMUNICATIVE DEMANDS (IproCo)

The theoretical framework we propose aims to bring together cognitive models as well as assumptions from communication pragmatics. It highlights the parallelism of individual problem-solving demands and communicative demands. IproCo includes task characteristics (concrete vs. abstract pictures) and situation characteristics (e.g. interactive vs. non-interactive...
communication) as important factors influencing the two demands. Further, it stresses the awareness of the task and situation-specific difficulties as a crucial prerequisite for a successful collaboration.

Although the framework is tailored to the collaborative picture-sorting task that we used, it can easily be applied to other collaborative tasks.

Individual Problem-Solving Demands

In our task, the individual problem-solving demands consist in visual search processes: In order to identify the relevant features (e.g. “position of the cat”) and feature values (e.g. “position of the cat: left vs. right”) that differ between the pictures, a serial visual search (e.g. Treisman & Gelade, 1980) has to occur. This requires the allocation of attention and is a time-consuming and error-prone process because focused attention is only a limited resource and cannot be directed to all details at the same time (Treisman & Schmidt, 1982). When observing a scene or looking at pictures, changes in objects or parts of objects are therefore often not detected – an effect called change blindness (e.g. Rensink, O'Regan, & Clark, 1997). This effect was found independently of the occurrence of saccades if attention was not allocated to the changing part of the picture (O'Regan, Deubel, Clark, & Rensink, 2000; Rensink et al., 1997). Such an effect was shown not only for change detection tasks, but also if the participants searched for differences in simultaneously presented pictures (Brunel & Ninio, 1997; Hansen, Spada, & Schneider, 2007; Scott-Brown, Baker, & Orbach, 2000). Change blindness occurs also in social settings (Simons & Levin, 1998) and has even been identified as a possible reason for inaccuracy in witnesses to crime (Davis, Loftus, Vanous, & Cucciare, 2007).

The visual salience of the differences, such as size, location, or color, was found to influence the detection of change (Smilek, Eastwood, & Merikle, 2000; Williams & Simons,
2000). However, differences in low-level factors such as visual salience are not the only
determinants for the allocation of attention. Rather, high-level factors such as familiarity or
semantic differences have been shown to highly influence the detection of changes (O’Regan et
al., 2000; Pearson & Schaefer, 2005; Rensink et al., 1997). Rensink and colleagues (Rensink et
al., 1997; O’Regan et al., 2000) used the degree of interest of a picture part as an
operationalization and found that changes are more likely to be detected if they occur on parts of
central interest than of marginal interest.

In both experiments reported here, we used two types of pictures that differed according
to their task characteristics, i.e. familiarity.

The concrete pictures contained familiar or known objects, whereas the abstract ones
depicted geometrical figures or patterns of figures (see Figure 1). The search for features in a
concrete picture resembles the process of finding differences in a so called ‘find the difference’
task, whereas searching for features in abstract tasks is more comparable to processes of pattern
recognition. Table 1 lists the task characteristics and demands that were assumed to vary
between concrete and abstract pictures.

To verify the differences in the task characteristics, we presented the pictures to 46
students and asked them to rate the pictures on a 5-point scale regarding their familiarity (see
Snodgrass & Vanderwart, 1980). The participants rated the concrete pictures ($M = 4.1; SD = .96$)
as significantly more familiar than the abstract pictures ($M = 2.1; SD = .90; t[45] = 11.35, p <
.001$).
Though, the prediction was not, that higher familiarity leads to a decrease of individual cognitive demands. We rather assumed familiarity to have an increasing effect because of the differences in salience of picture parts: In concrete pictures, there are known objects and less known parts. These differences imply also differences in the degree of interest. In contrast, our abstract pictures do not contain as much differences in the degree of interest. Parts with a higher degree of interests are known to draw attention, impeding the switch of attention to picture parts with an inferior degree of interest (see Rensink, 2002). The attention ‘sticks’ on the familiar parts or objects. In our task, a systematic scanning of all picture parts is required for successfully identifying the relevant features. Therefore, we assume higher difficulties in detecting the relevant features differing between concrete pictures (as familiarity brings along more important differences in the degree of interest) and make the counter-intuitive prediction of high individual problem-solving demands for concrete pictures, i.e. more errors in detecting the relevant features.

Communicative Demands

The communicative demands consist for the speaker in providing an understandable description of the relevant features. Hence, both collaborators have to assure mutual understanding for the features (e.g. “big eye circle”) and the feature values (e.g. “eye circle looking up vs. down”) by installing referential identity (Clark & Brennan, 1991): Speaker and addressee have to assign the same meaning to a name or an expression and to match it to the same feature and values.

To promote referential identity, a speaker can engage in audience design (e.g. Clark & Murphy, 1982) by tailoring his or her utterances to the knowledge or the perspective of the specific addressee. However, it is not always obvious for the speaker which description can
permit more than one meaning. She has to learn through the reactions of the addressee when
audience design is necessary (Horton & Gerrig, 2002) and then has to be aware of such task and
description characteristics.

One task characteristic determining the communicative demands is – again – the
familiarity of an object: Familiar or concrete objects with a high degree of linguistic codability
are easier to describe (e.g. Barry, Morrison & Ellis, 1997; Bortfeld et al., 2001; Hupet et al.,
1991; Oldfield & Wingfield, 1965) and the range of possible expressions is much smaller than
for unknown or abstract objects (Bertholet & Spada, 2005; Horton & Gerrig, 2002). With
unfamiliar objects, the speaker has to invent terms and to introduce them to the addressee in
order to establish referential identity (Isaacs & Clark, 1987); the collaborators have more
difficulties in negotiating a conceptual pact (Brennan & Clark, 1996). Therefore, we expect
abstract pictures to be more difficult to describe (see Table 1).

One situation characteristic determining the communicative demands is the mode of
communication. While interactive communication is the main setting for language use (e.g.
Pickering & Garrod, 2004), there are also settings involving less interactive communication or
settings without feedback options such as lectures and speeches or written communication (we
will use the term non-interactive to refer to less-interactive settings). Non-interactive settings do
not provide the speakers with the opportunities to learn about the need for audience design, as
the speaker does not receive feedback and cannot be certain of having been understood by the
addressee: Grounding (Clark & Brennan, 1991; Clark & Schaefer, 1987) is not possible and
more communicative effort is needed to overcome the hindering influences (Bromme, Hesse, &
Spada, 2005; Münzer & Borg, 2007). Non-interactive settings also affect the addressee’s
possibility to point out missing information about non-detected features and eliminate the possibility of identifying non-shared distributed knowledge of features.

The situation is further characterized by the communication media used (Clark & Brennan, 1991). In contrast to face-to-face settings, participants in net-based communication do not share the same physical environment (no copresence) and cannot use pointing gestures or visual gazes to specify their referents. Therefore, a first possibility to support the collaborators is to enhance the technical environment and to provide them with a kind of copresence. For example, shared workspaces offer visual information of task objects and participants’ actions (e.g. Tang, 1991; Whittaker et al., 1993). They can be used to store information, express ideas, and mediate interactions (Tang, 1991). The simplest version of electronic shared workspaces is shared whiteboards on which remote collaborators can draw or write. Yet, this simple form allows gesturing and enables an easy identification of objects and locations as well as depiction of spatial relations (Whittaker et al., 1993) for drawing applications. Information entered into a shared text-editor can be referred to at any time of the collaboration process and allows for review, in terms of Clark and Brennan (1991).

To facilitate net-based collaborative learning or work a number of tools have been developed; many including sophisticated shared representations or workspaces (e.g. Bell, 1997; Erickson & Kellogg, 2003; Roseman & Greenberg, 1996; Suthers & Weiner, 1995). However, in our first experiment we used very simple shared applications, because our aim was to scrutinize basic individual problem-solving and communication processes and to explore the influences of shared applications on these processes.

A second starting point – which we used in our second experiment – is to improve the awareness of the need for audience design. This can be done with the help of instructional measures.
EXPERIMENT 1

This first experiment aimed to answer the following research questions, in addition to the more general goal of gaining insights for the design of effective instructional support measures:

1. Do the collaboration process and the outcome of the collaboration differ with the task characteristics? For concrete pictures, we expected the collaboration process to show more difficulties due to the individual problem-solving demands and for abstract pictures, to show more difficulties due to the communicative demands. Regarding the outcome, either both demands could have the same impact, or one of the two demands could have more impact. In the first case, there would be no difference in the outcome between both types of task, whereas the outcome should differ in the latter.

2. Do process and outcome differ between interactive and restricted communication conditions (situation characteristics)? We expected fewer errors in the collaboration process and better results in interactive conditions.

3. Does the use of a shared application have an impact on process and outcome? We expected shared applications to support the collaboration process and lead to better results.

Method

Participants

120 students (60 dyads) from the University of Freiburg, Germany participated in the study; 62 of them were male and 58 female. Students of psychology were excluded. The participants had an average age of 23.35 years ($SD = 3.7$; range = 20 to 37 years). All participants were German native speakers and had normal color vision. The experiment lasted for 60 to 90 minutes and each person received eight Euros for taking part. The participants were
randomly grouped into dyads and assigned the role of speaker or addressee. Participants did not know each other prior to the study.

**Design and procedure**

A 2x3 factor design was implemented, with the mode of communication (interactive vs. restricted communication) and the availability of a shared application (pictorial, textual, or none) as between-subject factors and with the type of pictures (concrete vs. abstract) as an additional within-subject factor.

Each dyad was assigned randomly to one of the six conditions. Each participant received instructions in the form of a manual explaining the use of the technical environment and the procedure. The dyad then performed a training task to familiarize themselves with the technical environment. They were instructed to complete the tasks as accurately and quickly as possible. During the experimental phase, each dyad was required to solve two concrete and two abstract tasks. All dyads received the tasks in the same sequence. The speaker described the pictures and their positions and the addressee placed the pictures on his or her display accordingly. After completing each task, the participants received feedback about the number of pictures they had placed in the correct position and were informed about the features that had been relevant in the task (see next section).

**Type of pictures.** The concrete tasks contained familiar objects, while the abstract tasks contained patterns or geometrical figures (see Table 1). The number of relevant features differed depending on the task (three and four features in concrete, four and five in abstract tasks); whereas the relevance of a feature was determined by the necessity to describe its value in order to identify a picture unambiguously. For one of the concrete tasks - the cat-task (see Figure 1) - these features were: the position of the cat, the position of the cat’s head, the direction of the
cat’s whiskers, and the form of the mouse’s tail (mouse on the bottom). An irrelevant feature was, for example, the position of the cat’s tail in addition to the position of the cat, as these features varied conjointly. However, if a speaker did choose to describe the position of the cat’s tail instead of describing the position of the cat, this was considered as a relevant feature.

*Mode of communication.* Depending on the condition, the addressee was either able to talk or not. In the latter case, the addressee did, however, have the possibility to send an audio warning signal to the speaker if there was any trouble in understanding. This enabled a comparison of interactive and restricted communication settings.

*Shared application.* Additionally, a pictorial or textual shared application was each presented in one interactive and one restricted communication condition. These applications influencing the situation characteristics were introduced as a second factor because their availability may lead to a kind of copresence (Clark & Brennan, 1991) and thus facilitate communication. The pictorial shared applications contained a picture of the respective task by way of example and had a pointing and drawing feature that enabled deictic gestures. The textual shared applications consisted of a shared text editor and allowed the relevant differences to be noted down. These notes were always present on both screens and could be looked at again during the collaboration process. In the conditions with a shared application, only the speaker was able to make drawings or write text, but these were also visible to the addressee. However, the speakers were not obliged to use the shared application to support their explanations, and were merely informed about the additional possibility of using it.

*Measures*

Three sets of data were collected to examine the process and the outcome of collaboration: (1) audio recordings of the verbal communication (process measures), (2)
performance data (performance measures), and (3) data of the shared applications (drawings or
text).

**Process measures.** To examine whether the postulated demands affected the
communication process, a coding scheme was developed consisting of categories for the
individual problem-solving demands and the communicative demands.

The first group included the *percentage of identified features*. This category was a rate
measure, as there was a different amount of relevant features depending on the task (three and
four features in concrete, four and five in abstract tasks). All other categories were frequency
measures, because they related to single picture descriptions (i.e. a set of the speaker’s utterances
describing one picture). The second category for the individual problem-solving demand – *the
number of “feature not mentioned” errors* – counted the number of picture descriptions in which
relevant features were not mentioned. Note that the number of picture descriptions could be
higher than nine, in case the speaker described some pictures more than once.

During coding, we made one important change to the categories for the communicative
demands: Initially, we counted all ambiguous descriptions reflecting failures to establish
referential identity. Then, we realized that this communicative error category was too broad and
in fact comprised two distinct types of errors: ambiguous descriptions regarding the feature’s
name or position and ambiguous descriptions regarding the frame of reference. A spatial
description can either be viewer-centered, object-centered, or environment-centered (see Levelt,
2003). Notably, in descriptions of concrete pictures, ambiguities may be expected regarding the
frame of reference used. A viewer-centered and an object-centered frame of reference can lead to
two opposite meanings (e.g. “the cat is looking to the left” – described from the cat’s or the
viewer’s perspective?). Therefore, we included two distinct categories, both reflecting failures to
establish referential identity: number of “name and position of feature” errors and number of “frame of reference” errors. Further, we included the number of repeated picture descriptions, which counted a set of the speaker’s utterances describing one picture. In restricted communication conditions, we counted the number of warning signals and in interactive conditions, we counted the number of clarification questions the addressee asked. We also counted the number of coordination activities, i.e. the number of times the speaker used meta-language to talk about the way in which to describe the features.

It should be noted that some categories are not independent of each other: Of course, if a speaker did not describe a relevant feature (individual problem-solving demand), then she could not commit errors in the description (communicative demand). The audio recordings of the verbal communication were analyzed using these categories. Ten percent of the verbal data was coded by a second rater to enable the calculation of inter-rater reliability. The consistency of the coding was medium to high, with mean intra-class correlation (e.g. Shrout & Fleiss, 1979) of .81, indicating that the coding scheme could easily be used.

Performance measures. The quality of the joint solution was measured by the number of pictures placed in the correct position and the time needed to complete the task.

Data of shared applications. In terms of the data gathered from the shared applications, the number of dyads that made use of them was counted. This data was then either coded as supporting the individual problem-solving demands (feature identification/ remembering the features), or as supporting the communicative demands (describing the feature values of each picture). In the first case, the speaker had marked or written down all features at the beginning of a task in order to make it clear what features generally differed between the pictures (e.g. “the mouse’s tail, the position of the cat…”). Whereas in the second case, the speaker had used the
shared application to mark or write down the *feature values* of each picture (e.g. “In picture 1, mouse’s tail in S-shape, cat on the left...”) during the description of the nine pictures.

**Results and Discussion**

*Process and performance data*

A multivariate analysis of variance (MANOVA) with repeated measures (for the factor type of pictures) was conducted to test the influence of the three factors (type of pictures, mode of communication, shared application) on the process and performance measures. There was an effect of the type of pictures ($F[9, 46] = 20.6, p < .01, \eta^2 = .80$), an effect of the communication mode ($F[9, 46] = 12.2, p < .01, \eta^2 = .71$), but no effect of the shared application. The lack of effect of the shared application may be due to the fact that not all dyads of the respective conditions used them, as their usage was not obligatory. Eighty percent of the dyads in the respective conditions used it for at least one task, but the number of dyads using the pictorial shared application differed notably from task to task and from condition to condition (varying between 10 and 90 percent). We will present the rate of usage in more detail in the next section (data of the shared application).

The MANOVA revealed also a significant interaction between type of pictures and mode of communication ($F[9, 46] = 2.9, p < .01, \eta^2 = .36$).

Table 2 contains means and standard deviations of all variables for the factors type of pictures and communication modes. As it had no significant effect, the factor shared application is not included in Table 2. In the following, we will describe the main effects for the factors type of pictures (second and third columns of Table 2) and mode of communication (last two columns), as well as for the significant interaction (not included in Table 2).
Main effects of “type of pictures”. With regard to the process measures, there were fewer features identified ($F[1, 54] = 3.4, p < .05, \eta^2 = .06$), more errors related to the individual problem-solving demands in tasks with concrete pictures (“feature not mentioned” errors; $F[1, 54] = 46.3, p < .01, \eta^2 = .46$), and fewer “name and position of feature” errors ($F[1, 54] = 22.3, p < .01, \eta^2 = .29$). The number of “frame of reference” errors was also higher in tasks with concrete pictures ($F[1, 54] = 5.2, p < .05, \eta^2 = .09$). Furthermore, there were fewer warning signals sent or questions asked ($F[1, 54] = 13.6, p < .01, \eta^2 = .20$) and fewer coordination activities carried out in concrete tasks ($F[1, 54] = 13.8, p < .01, \eta^2 = .20$). As expected, addressees were more likely to ask for additional descriptions or clarifications with abstract tasks (see also Horton & Gerrig, 2002) and the speakers were more likely to explain the process (meta-communication; cf. Hancock & Dunham, 2001). However, more repeated picture descriptions were made in concrete tasks ($F[1, 54] = 16.9, p < .01, \eta^2 = .24$). It can be assumed that this was related to the higher amount of not identified features, the higher number of “feature not mentioned” errors, as well as the higher number of “frame of reference” errors: Both the detection of overlooked features as well as the recognition of misunderstandings regarding the frame of reference may bring about the necessity to describe a picture more than once.

With regard to the performance measures, Table 2 displays the percentage of correctly placed pictures, as the maximum number is different for the factors type of pictures (maximum of 18 pictures for 2 x 9 concrete and 2 x 9 abstract pictures, respectively) and mode of communication (maximum of 4 x 9 pictures). Fewer pictures were placed in the correct position in tasks with concrete pictures ($F[1, 54] = 63.5, p < .01, \eta^2 = .54$), although it took less time to complete them ($F[1, 54] = 105.8, p < .01, \eta^2 = .66$). This may seem to be a time-accuracy trade-off, but the result can also be interpreted in light of the different demands: The individual
problem-solving demands of feature detection were indeed more important in concrete tasks, leading to fewer correctly placed pictures. In abstract tasks, the speakers had to make additional effort to engage in audience design; as a consequence, more time was spent in completing the tasks. This can be interpreted as an illusion of simplicity effect (Nickerson, 1999) for concrete pictures: The familiarity of the concrete pictures led to the deceptive impression that the differences were easy to detect and the pictures easy to describe. Therefore, the participants took less time for search and description and committed more errors.

To scrutinize this effect, we asked 46 students about their impressions of each picture from Experiment 1: They had to rate on a 5-point scale their impression of how easy it would be to find the differences and how easy to describe them. The participants attributed significantly fewer difficulties to finding the features for the concrete \( (M = 2.0; SD = .79) \) than for the abstract pictures \( (M = 3.9; SD = .79; F[1] = 140.1, p < .01, \eta^2 = .76) \). Comparable results were found for describing the features (concrete: \( M = 1.9; SD = .75 \); abstract: \( M = 3.7; SD = .87; F[1] = 217.9, p < .01, \eta^2 = .83) \).

**Main effects of “mode of communication”**. Starting again with the process measures, there were more features identified \( (F[1, 54] = 5.2, p < .05, \eta^2 = .09) \), fewer “name and position of feature” errors made \( (F[1, 54] = 3.1, p < .05, \eta^2 = .06) \), more questions asked \( (F[1, 54] = 46.5, p < .01, \eta^2 = .46) \), and more coordination activities carried out \( (F[1, 54] = 9.9, p < .01, \eta^2 = .16) \) in interactive conditions.

With respect to the performance data, the dyads in interactive conditions placed approximately one picture more in the correct position (ten percent points more than in abstract tasks; \( F[1, 54] = 3.1, p < .05, \eta^2 = .11 \)) and took roughly 100 seconds less to complete the task \( (F[1, 54] = 6.8, p < .01, \eta^2 = .06) \). This is in line with previous findings showing the additional
difficulties of restricted communication settings (e.g. Clark & Krych, 2004; Krauss & Weinheimer, 1966; Schober, 1993). 

*Interactions “type of pictures” and “mode of communication”.* In concrete tasks, the number of “feature not mentioned” errors was particularly high in conditions with restricted communication ($M_{\text{concrete\_restricted}} = 11.1; M_{\text{concrete\_interactive}} = 6.9; M_{\text{abstract\_restricted}} = 3.1; M_{\text{abstract\_interactive}} = 3.6; F[1, 54] = 8.1, p < .01, \eta^2 = .13$). Further, there were fewer repeated picture descriptions in interactive conditions for the concrete tasks ($M_{\text{concrete\_restricted}} = 21.0; M_{\text{concrete\_interactive}} = 6.9; M_{\text{abstract\_restricted}} = 19.5; M_{\text{abstract\_interactive}} = 19.8; F[1, 54] = 5.6, p < .05, \eta^2 = .09$). In abstract tasks, addressees in interactive conditions asked more questions ($M_{\text{concrete\_restricted}} = 3.3; M_{\text{concrete\_interactive}} = 7.1; M_{\text{abstract\_restricted}} = 3.7; M_{\text{abstract\_interactive}} = 11.1; F[1, 54] = 8.8, p < .01, \eta^2 = .14$).

**Data of shared applications.**

The use of the shared application – if available – was not obligatory. As the number of dyads having used the shared application in one task was quite small (varying between 6 and 14 dyads out of 20), the following closer look at the rate, the benefit, and the type of use should be treated with caution.

Overall, the rate of use of the pictorial shared application was higher than the rate of use of the textual shared application: If available, eighty percent of the speakers used the pictorial shared application for at least one task, whereas 50% of the speakers used the textual application for at least one task. However, the number of dyads using the shared application differed notably from task to task: More dyads used the pictorial shared application in abstract tasks (62.5%) compared to concrete tasks (40%) and the rate of the textual shared application was higher in concrete tasks (55%) compared to abstract tasks (37.5%). The dyads seem to have used the
possibility of indicative gestures for establishing referential identity predominantly for the abstract tasks (see Tang, 1991). As it is easier to write down features with already existing and known names, the dyads used the textual shared application more often with concrete tasks.

The rate of use of the pictorial shared application was comparable between interactive (50%) and restricted communication conditions (52.5%), whereas the rate of use of the textual shared application was higher for conditions with restricted communication (55%) compared to interactive conditions (37.5%). It can be supposed, that speakers in conditions with restricted communication were more willing to invest the additional effort of typing feature descriptions into the shared text editor to ensure a better understanding (Schober, 1993).

To inspect the possible benefit of using the shared application, a – still only descriptive – comparison of dyads that used vs. dyads that did not use the shared application may be helpful: When using the shared application, the dyads took more time to complete the task ($M_{sa\_used} = 276; M_{sa\_notused} = 236$). Was this disadvantage worthwhile? For the pictorial shared application, dyads in interactive conditions committed fewer feature not mentioned errors ($M_{p\_sa\_used\_inter} = 1.8; M_{p\_sa\_notused\_inter} = 2.4$) and fewer name and position of feature errors ($M_{p\_sa\_used\_inter} = 0.2; M_{p\_sa\_notused\_inter} = 0.6$) if they made use of the application. For the textual shared application, dyads in conditions with restricted communication committed fewer feature not mentioned errors ($M_{t\_sa\_used\_non\_inter} = 3.8; M_{t\_sa\_notused\_non\_inter} = 5.0$) and could therefore place more pictures on the correct position ($M_{t\_sa\_used\_non\_inter} = 7.0; M_{t\_sa\_notused\_non\_inter} = 6.5$) if they used the textual shared application. To summarize: Interactive dyads profited best from pictorial, restricted communication dyads profited best from textual shared applications.

The data of the shared application (the speaker’s drawing or text) can be used to further analyze the impact of the problem-solving and communicative demands. For this purpose, the
data was either coded as *used to support the individual problem-solving demands*, if the speaker used the shared application to support feature identification (What are the *features* that differ between the pictures?), or as used *to support the communicative demands*, if the speaker used the shared application to describe the feature value of each single picture (What are the feature *values* of picture 1, picture 2, …?). Figure 2 exemplarily shows data of the shared pictorial application for both types of usage.

To further illustrate the usage of the shared application, consider the following descriptions, each corresponding to one of the two pictures in Figure 2. The speakers drew into the shared application while pronouncing the underlined words.

(1) Coded as *used to support the individual problem-solving demands* (What are the *features* that differ between the pictures?):

S: You have to consider the red fields, the broken and the continuous lines. First picture: Outer red field between nine and eleven. […]

(2) Coded as *used to support the communicative demands* (What are the feature *values* of picture 1, picture 2, …?):

S: I’m starting with the first one. Well, there is a circle and this line here, only this line is black… and then this line…this… this… and this. And this circle thing here. And what’s red, it is this box here, this, and this – they are red.

In both examples, the pictorial shared application was used to establish referential identity with less effort. This facilitation becomes clear when comparing the above examples with the following description made by a speaker who had no shared application available:
Regarding the type of use, both shared applications were mostly used to support the communicative demands, i.e. the speaker used the shared application to describe the feature value of each single picture instead of using it to support feature identification. We will discuss practical implications of these findings for the supply of shared applications in the overall discussion.

Implications for designing instructional support measures

As one of our aims was to develop effective instructional support measures based on the findings of first experiment, what did we learn that can help in the development of these measures? First, the results support the theoretical assumption of the IproCo framework: Individual problem-solving and communicative demands both have an impact on the collaboration process and outcome. As predicted, they are influenced by the familiarity of the task characteristics, with higher individual problem-solving demands arising in concrete tasks and higher communicative demands arising in abstract tasks. They are further influenced by the situation characteristics mode of communication (interactive/ restricted), with higher demands in restricted communication conditions. Moreover, the process analysis helped to identify a second task characteristic: the possibility for different frames of reference. Thus, a sustainable support should aim at reducing the number of errors related to both demands by improving the awareness: For concrete pictures, it should be emphasized to the participants that they should
carefully search for all feature differences and establish a mutual frame of reference. For abstract pictures, the need to define understandable names for the features should be pointed out.

EXPERIMENT 2

Several approaches attempt to foster collaboration by structuring the problem-solving and communication process. *Scripted cooperation* (O’Donnell & Dansereau, 1992) is a prominent technique that has also been transferred to computer settings (e.g. Baker & Lund, 1997). It aims at optimizing the interaction process by sequencing it into different phases, defining roles, and assigning them to the collaborative learners. Typically, they are embedded into computer-based learning environments and guide the collaborators in a step-by-step fashion through different activities. A second approach provides the participants with a *model collaboration* prior to the actual collaboration (Rummel & Spada, 2005). While observing such a model of a dyad collaborating, people should engage in meta-cognitive activities that promote learning (e.g. Bandura, 1977; VanLehn, 1996).

Both approaches have been shown to improve the collaboration, but still entail certain disadvantages: The extensive use of collaboration scripts can disturb natural interaction or cognitive processes and lead to motivational losses (overscripting; see Dillenbourg, 2002). Depending on the amount of information included, the persons watching the model collaboration might have difficulties in extracting and remembering all relevant points. To combine the advantages of the two approaches, we developed instructional support measures integrating both. Following the approach of Rummel and Spada (2005), we developed a model collaboration showing a successfully collaborating dyad for the collaborative picture-sorting task. It was presented to speaker and addressee individually as an on-screen video with audio instructions prior to the collaboration. In addition, a collaboration script reminding the collaborators of what
they had just learned was provided during the collaboration. This combination should promote
effective collaboration, as we expect the model collaboration video to have positive effects on
the participants’ attention and motivation (Bandura, 1977) and the collaboration script to
enhance the memory for the relevant information.

The instructional measures (model instruction and script) were aimed at augmenting the
awareness for the task characteristics necessitating audience design and to transform the parallel,
dual task structure into a sequential one: (1) searching for features, (2) finding feature’s names as
to describe the relevant features in an understandable way, and (3) describing the single pictures.
The measures contained two levels of support: level 1, corresponding to the subtask “searching
for features” and thereby supporting the individual problem-solving demands, and level 2,
corresponding to the subtask “finding names for the features” and thus supporting the
communicative demands. Each support level contained hints for dealing with the respective
demand and, moreover, each level introduced one subtask to force the speakers to follow the
sequential task structure: marking the features in an individual picture editor (level 1) and writing
the features’ names into an individual text editor (level 2). It should be noted, that these editors
were no shared applications as in Experiment 1, but individual editors. In contrast to Experiment
1, they were not introduced to be used as additional communication channel but to guarantee the
intended sequential structure of the task with one demand after another to be dealt with: Marking
the features in the individual picture editor should help to identify all relevant features and
values. Writing down the features’ names should help to formulate appropriate descriptions and
– combined with hints for the description – to be sensitized for pitfalls hindering audience
design.
The results of the first experiment showed restricted communication to be a very difficult undertaking. In order to test for effects of support, we included interactive as well as non-interactive conditions in Experiment 2 and provided both with instructional support measures. This time, we established a more rigorous non-interactive condition: Speaker and addressee in non-interactive conditions performed the tasks not at the same time, but rather one after another. The speaker recorded his or her description and the audio recording was given to the addressee later on. However, we provided both interactivity conditions with the same support measures, as we focused on the awareness for task characteristics.

The second experiment aimed at answering the following research questions:

(1) Do the instructional support measures improve collaboration process and outcome? We expected to reveal an impact of the instructional support measures in both process and outcome, with complete support (level 1 and 2) being most beneficial.

(2) Is there a difference in the impact of the measures in non-interactive compared to interactive communication conditions? We expected our support measures to be helpful for both interactive and non-interactive conditions.

Method

Participants

Ninety-six students (48 dyads) of the University of Freiburg, Germany participated in the study. Students of psychology were excluded. Thirty-six of the participants were male and 60 were female. The participants had an average age of 24.15 years ($SD = 4.4$, range = 18 to 48) and all were German native speakers. The experiment lasted for 90 to 120 minutes and each participant received 15 Euros for taking part. The participants were randomly grouped into
dyads, assigned the role of speaker and addressee, and randomly assigned to one of the six conditions. Participants did not know each other prior to the study.

Design and procedure

A 2x3 factor design was used, with the amount of support (no support/ level 1 support/ level 1 and 2 support) and the mode of communication (interactive/ non-interactive) as between-subject factors. Again, the type of pictures (concrete/ abstract) was varied as an additional within-subject factor.

Prior to the collaboration, the participants individually received instructions including technical advice in the form of an on-screen video. After watching the on-screen video, the dyads performed a training task in order to familiarize themselves with the technical environment and the subtasks. During the experimental phase, each dyad was required to solve two concrete and two abstract tasks but in different sequences: To control for sequence effects, four different task sequences were given. However, as an ANOVA revealed no effect of the task order, this factor will not be taken into account in the subsequent analyses. As in the first experiment, 9 pictures had to be described and positioned. Yet, this time, the addressee had 16 pictures per task to choose from on his or her display. The speaker still had 9 target pictures but saw 7 additional pictures on his or her display. As more pictures were presented, the difficulty of choosing the correct picture from the remaining ones was therefore still present for the 8th and 9th pictures.

Amount of support. Depending on the condition, the on-screen video included only technical advice (no support conditions) or additionally contained the support of the problem-solving demands (level 1 support) or of both demands (level 1 and 2 support). Level 1 support contained hints for feature search (e.g. pointing out the counter-intuitive higher difficulty of feature detection in concrete task) and introduced the first individual subtask (marking the
differences in the individual picture editor). Moreover, level 1 and 2 support included hints for dealing with the communicative demands (e.g. demonstrating the utility of explicit grounding of the features and of specifying the frame of reference used) presented as model collaboration, by displaying textboxes with the model speaker’s descriptions. Additionally, the second individual subtask was introduced (writing the features’ names in the individual text editor). By making the division into individual and collaborative subtasks explicit, the instructional support measures were meant to overcome the collaborators’ tendency to solely engage in joint activities and forget about individual work phases (Hermann, Rummel & Spada, 2001). The parallelism of the two demands should further be fostered by the support measures.

*Mode of communication.* In the non-interactive conditions, the speaker and addressee performed the tasks not at the same time, but rather one after another. The speaker had a recording device on the display and was able to start and stop recording the explanations to the addressee as desired. A microphone was positioned on the table next to the monitor. The description of each speaker was randomly assigned to one addressee, who later arranged the pictures according to the recorded descriptions. The addressee had an audio-player device on the display and was able to start, stop and rewind the recording of the speaker’s descriptions.

However, speakers and addressees in both the non-interactive and interactive conditions were told to proceed as accurately and quickly as possible.

*Measures*

Two sets of data were collected to examine the process and the outcome of collaboration: (1) audio recordings of the verbal communication (process measures, time periods) and (2) performance data (performance measures).
Process measures. The coding scheme used to analyze the audio recordings of the verbal communication data consisted of most categories used in the first experiment. Three new categories replaced three categories used in the first study (number of repeated picture descriptions, number of warning signals or clarification questions asked, number of coordination activities), which could not be used to analyze the process of our new, non-interactive conditions: The percentage of features revealed before description indicated how many relevant features the speaker revealed before starting the description of the first picture. We used a proportional measure, as there were five relevant features in concrete and four relevant features in abstract tasks. The number of irrelevant features counted the number of times the speaker gave descriptions of irrelevant features. The number of complicated descriptions counted the number of times the speaker gave complicated and laborious descriptions of a feature. These complicated descriptions were for example characterized by disfluencies as repeats, restarts, fillers or editing expressions (see Bortfeld et al., 2001) as well as a high number of words Again, ten percent of the verbal data was coded by a second rater to enable the calculation of inter-rater reliability. The consistency of the coding was medium to high, with mean intra-class correlation (e.g. Shrout & Fleiss, 1979) of .85, indicating that the coding scheme could easily be used.

Time periods. Additionally, the amount of time used before starting the description of the single pictures was measured, allowing to identify two time periods: an individual phase in which speaker and addressee searched for features and labeled the features individually in silence (individual time), and an information period (information time). During this second period, the speaker informed the addressee about the relevant features he found to differ between the pictures, about the labels he chose for them, and about the frame of reference he would use to describe the pictures. This information period is a necessary process for achieving mutual
understanding and corresponds to a part of the grounding process. According to Clark and Schaefer (1987), the grounding process consists of a presentation and an acceptance phase. In asynchronous conditions, the information phase corresponded to the presentation phase, whereas in synchronous conditions, it also included the acceptance phase, during which the addressee gives evidence of understanding or misunderstanding.

In conditions without support, there was no explicit direction for structuring the process into these time periods. And in level 1 support conditions, only the individual period was instructed. However, we discerned the individual time before the speaker started the description and the information time for all dyads, as we believed them to be necessary steps of the problem-solving and communication process and some speakers in no or level 1 support conditions made spontaneously use of them. Speakers that did not use an individual time period performed feature search and description processes simultaneously.

**Performance measures.** As in the first experiment, the performance measures included the number of pictures placed in the correct position and the time needed for picture description and positioning. Thus, the total time on task was composed of the individual time, the information time, and the time needed for picture description and positioning.

**Results and Discussion**

We computed a MANOVA with repeated measures (for the factor type of pictures) in order to test the influence of the three factors on process and performance measures. There was an effect of the type of pictures ($F[11, 32] = 12.8, \ p < .01, \ \eta^2 = .81$), an effect of the amount of support ($F[22, 66] = 2.8, \ p < .01, \ \eta^2 = .48$) and an effect of the mode of communication ($F[11, 32] = 9.5, \ p < .01, \ \eta^2 = .77$). There were no significant interactions. Table 3 contains means and standard deviations for all variables. In the following, we will describe the main effects for the
factors type of pictures (second and third columns of Table 3), amount of support (columns four, five, and six of Table 3), and mode of communication (last two columns).

Main effects of “type of pictures”

Starting with the process measures, there were fewer identified features \( F[1, 42] = 32.0, p < .01, \eta^2 = .43 \), a higher number of “feature not mentioned” errors \( F[1, 42] = 80.6, p < .01, \eta^2 = .66 \), fewer features revealed before description \( F[1, 42] = 17.2, p < .01, \eta^2 = .29 \), a higher number of “frame of reference” errors \( F[1, 42] = 14.7, p < .01, \eta^2 = .26 \) and of irrelevant features \( F[1, 42] = 28.4, p < .05, \eta^2 = .40 \) in concrete tasks. This is in line with the results of the first experiment. On the other hand, the number of “name and position of feature” errors \( F[1, 42] = 22.2, p < .01, \eta^2 = .35 \) as well as the number of complicated descriptions \( F[1, 42] = 21.8, p < .01, \eta^2 = .34 \) arose more often in abstract tasks.

With regard to the time periods, there was more individual \( F[1, 42] = 21.7, p < .01, \eta^2 = .34 \) and also more information time used \( F[1, 42] = 7.7, p < .01, \eta^2 = .16 \) in abstract tasks.

Regarding the performance data, there were more pictures placed in the correct position for abstract than for concrete pictures \( F[1, 42] = 39.1, p < .01, \eta^2 = .48 \). This result is in line with the findings of Experiment 1. However, there was no significant difference in the time needed for description and positioning of the pictures.

Main effects of “amount of support”

Starting again with the process measures, the means show that fewer problems and errors occurred in conditions with support. Nevertheless, there were significant differences only in the number of “feature not mentioned” errors \( F[2, 42] = 5.2, p < .05, \eta^2 = .20 \) and in the percentage of features revealed before description \( F[2, 42] = 13.9, p < .01, \eta^2 = .40 \). As expected, both
problems occurred mostly in conditions without support: The speakers searched for the features and started with the description in parallel, leading to more errors.

With regard to the time periods, the dyads with complete support separated feature search (individual time), grounding of features and values (information time), and description of single pictures (time needed for picture description and positioning), as intended by the support measures: These dyads took the highest amount of individual time, followed by dyads with support of the individual problem-solving demands and finally the dyads without support ($F[2, 42] = 41.1, p < .01, \eta^2 = .66$). The information time of dyads with complete support was three times longer than of dyads in the other two conditions ($F[2, 42] = 8.7, p < .01, \eta^2 = .29$). In this information period, the speakers listed the features they had found, presented the names for features and values, and defined the frame of reference for their descriptions. In conditions without support or only with level 1 support, the lack of explicit information brought problems and irritations. The following dialog of a dyad without support illustrates this (concrete task “cat”, first picture; see Figure 2).

(4) S: The cat’s head is directed upwards to the left.
A: Head tilted to the left.
S: No, wait. Not tilted to the left! Directed upwards to the left.
A: But that’s tilted to the left.

The performance measures, however, did not reflect the impact of the measures unanimously: Dyads without support took more time for description and positioning of the pictures than dyads with support of the individual demands as well as dyads with complete support. Dyads with complete support were faster than dyads in the other conditions ($F[2, 42] = \ldots$)
2.5, \( p < .05, \eta^2 = .11 \)). Yet, there was no significant difference in the number of correctly placed pictures between the conditions with different amounts of support.

**Main effects of “mode of communication”**

With regard to the process measures, the means of problems and errors during communication differed mostly in the expected way, with fewer problems and errors occurring in interactive conditions. Still, there were only significant differences in the percentage of identified features \( (F[1, 42] = 49.4, p < .01, \eta^2 = .54) \), in the number of “feature not mentioned” errors \( (F[1, 42] = 17.9, p < .01, \eta^2 = .30) \), and in the percentage of features revealed before description \( (F[1, 42] = 9.0, p < .01, \eta^2 = .18) \). As expected, all three errors were made more often in dyads with non-interactive modes of communication.

With regard to the time periods, dyads in non-interactive conditions took more individual time \( (F[1, 42] = 13.5, p < .01, \eta^2 = .24) \), but less information time \( (F[1, 42] = 6.8, p < .05, \eta^2 = .14) \) than dyads in interactive conditions.

With respect to the performance measures, dyads with an interactive mode of communication placed more pictures in the correct position \( (F[1, 42] = 72.6, p < .01, \eta^2 = .64) \) than dyads with a non-interactive mode of communication.

**OVERALL DISCUSSION**

**Theoretical Implications - The Impact of the IproCo Framework**

Our first aim was to contribute to a better understanding of the processes underlying remote collaboration. We proposed the IproCo Framework, which includes assumptions from cognitive models and communication pragmatics, as well as factors influencing both individual problem-solving and communicative processes (task and situation characteristics). In both experiments, we used two types of pictures (concrete and abstract), which differed in terms of
familiarity and possibility for different frames of reference (task characteristics). Consequently, they should also vary in the amount of individual problem-solving and communicative demands. The results of both experiments confirmed the expected effects on process and outcome of the collaboration: In concrete tasks, the feature search (individual problem-solving demands) was more error-prone, and more “frame of reference” errors (one aspect influencing communicative demands) occurred. In abstract tasks, the communicative demands proved to be predominant in the collaboration process. The outcome of the collaboration may seem to be a time-accuracy trade-off, as there were fewer pictures placed in the correct position in concrete tasks and also less time used. Indeed, we interpret this in the light of the different demands: As expected, the features were more difficult to find in concrete tasks, more features were overlooked, and fewer pictures were placed correctly. In abstract tasks, the speakers spent more time for the descriptions due to the additional effort for audience design (Horton & Gerrig, 2002). We interpreted this as an illusion of simplicity (Nickerson, 1999) that occurred for concrete pictures: The familiarity of the concrete pictures led to the deceptive impression that the differences were easy to detect and the pictures easy to describe. Therefore, the participants took less time for search and description and committed more errors. Interactivity is one example of situation characteristics, which were also expected to influence individual problem-solving and communicative processes. In both experiments, we compared conditions enabling normal bidirectional, interactive communication to conditions enabling no or only restricted feedback from the addressee, i.e. restricted communication or non-interactive conditions. This allowed an examination of the individual achievement of the speaker without interaction and feedback effects in non-interactive or restricted conditions. Furthermore, the collaboration and problem-solving processes could be compared in order to examine whether speakers in non-interactive
conditions could compensate for the lack of feedback. Not surprisingly, dyads in interactive conditions had better results than dyads in restricted or non-interactive conditions. As the analysis of the process data showed, this was mostly due to the fact that, working together, two people identified more features. Furthermore, they had the possibility to speak about problems and the planned procedure. The illusion of simplicity discussed above was even more important in non-interactive conditions, since the number of features that the speaker did not identify at all was especially high in concrete tasks. To summarize, less interactive or non-interactive settings amplified the predominant demands – individual problem-solving in concrete and communicative demands in abstract tasks (see Table 1).

The IproCo Framework was shown to foster the understanding of collaboration process and outcome. So far, it has mostly focused on the role of speaker. Further research has to be conducted in order to investigate the processes and influencing factors for the role of addressee.

Practical Implications

*Shared applications*

Our second aim was to provide effective support to the collaborators. In the first experiment, we tested technical support measures (the availability of a pictorial or textual shared application) to improve the situation and enhance the possibilities for grounding. As the usage was not obligatory, only few dyads used the shared application (varying between 6 and 14 dyads out of 20). Therefore, no reliable effect could be found for the availability of a shared application.

Why did not all dyads use the shared application, if available? Many systems for remote collaboration provide shared applications (e.g. Bell, 1997; Erickson & Kellogg, 2003; Roseman & Greenberg, 1996; Suthers & Weiner, 1995). However, communication via shared applications
holds additional costs and has not always been shown to improve the communication process (e.g., Ertl, Fischer & Mandl, 2006; Dillenbourg & Traum, 1999; Whittaker et al., 1993). To learn about the role of shared applications for supporting individual problem-solving and communicative demands, we inspected the rate and the benefit of use on a descriptive level. Overall, the costs of usage seemed lower for pictorial shared applications, as more dyads used the pictorial compared to the textual shared application. In dependence of the type of task, the rate of use of pictorial and shared application differed: More dyads used the pictorial shared application in abstract tasks, whereas more dyads used the textual shared application in concrete tasks. The possibility for indicative gestures to support the establishment of referential identity—a function inherent to the pictorial shared application only—seemed to have been perceived as helpful particularly in abstract tasks. On the other hand, the costs of typing descriptions in the textual shared application had been lower for concrete tasks. Especially non-interactive dyads were willing to invest this additional effort, as they used the textual shared application more often.

A comparison of dyads that did vs. did not use the shared application revealed a benefit of use depending on the condition: Dyads in interactive conditions benefited from the use of the pictorial shared application, as they committed fewer feature not mentioned errors and fewer name and position of feature errors if they made use of the application. Dyads in non-interactive conditions benefited from the use of the textual shared application, as they committed fewer feature not mentioned errors and could therefore place more pictures on the correct position.

We presented examples from the verbal communication data and the data of the shared applications (drawings and text) in order to illustrate the benefit of the shared applications for establishing mutual understanding and to show the main type of use: Both shared applications
Supporting Remote Collaborative Problem-Solving

were mostly used to describe the feature value of each single picture (communicative demand) instead of using them to support the individual problem-solving demands, i.e. by supporting feature identification or using them as a kind of external group memory (Dillenbourg & Traum, 1999).

Thus, for the design of systems for remote collaboration, our results imply tailoring the shared applications to support the communicative demands of a task. This does not necessarily mean providing highly task specific shared applications (e.g. Ertl et al., 2006) but the situation and task characteristics that influence the communicative demands should be closely taken into account. In the following, we exemplarily show how such an analysis using the IproCo-Framework can help predicting the need for the availability of shared applications.

When to provide which kind of shared application?

As stated above, the IproCo Framework can help to clarify task and situation characteristics, as well as demands on the collaborators and help to decide if or what kind of shared applications should be provided. To exemplify such an analysis, we used IproCo to reinterpret the findings from Whittaker and colleagues (1993), who presented results from three studies with different tasks solved by remote dyads. In these tasks, the dyads either communicated only with an audio link or with an additional shared workspace. In terms of the IproCo framework, they varied the situation characteristic of shared physical environment with a kind of physical copresence through the availability of the shared workspace. To explain the differences in terms of the benefit of shared workspaces in the three tasks, the researchers also used the concept of demands. However, the terms were not well defined. The three tasks were described as follows: an “undemanding” text task, in which the dyad had to generate a list, a “demanding” text-editing task, in which the collaborators had to co-annotate a shared text, and a

http://mc.manuscriptcentral.com/acp
graphical design task, with two sets of unshared abstract objects that had to be arranged altogether in one plan. The results showed benefits of the shared workspace only in the third task, and after some practice for the second task.

In terms of the IproCo framework, the individual problem-solving demands were very low in the first task (remembering appropriate words). In the second task, the individual problem-solving demands consisted in text understanding and visual search for specific words or passages. In the third task, by contrast, the collaborators faced multiple individual problem-solving demands, for example those of visual search, object recognition, mental rotation, and memorizing positions of objects.

With regard to the communicative demands, the first task was quite undemanding, as the dyads were composed of members from the same work group and talked about familiar topics with perfect common ground. The second task required the establishment of referential identity for the text passage they were currently speaking of. The third task, again, required the most communicational effort: Audience design was crucial because the participants had to describe unfamiliar, abstract objects.

As both demands were low in the first task, the additional shared workspace did not show any effect. In the second task, the visual channel enabled pointing and helped to install referential identity, but this added value was not obvious from the start. However, in the third task, the shared workspaces, allowing for physical copresence, had an influence on both kinds of demands: They supported cognitive processes such as mental rotation and memory processes and also communicative demands (see results of our first experiment).

**Instructional Support Measures**
In the second experiment, we did no longer use shared applications to support remote collaboration but aimed to enhance the awareness of the need for audience design and used instructional support measures combining a collaboration model and an analogical collaboration script. Two levels of support were realized, the first one sustaining the problem-solving demands and the second the communicative demands. The analysis of the verbal data indeed showed the intended effect, with fewer problems and errors in conditions with support. Unfortunately, the performance did not reflect the impact of the measures across the board: The complete support conditions did take less time for the description and positioning of the pictures, but the total time on task (including individual and grounding time) was higher and there were no significant differences between the numbers of correctly placed pictures for the three support conditions.

This may be due to the high variance in the number of detected features, a measure which has a crucial influence on the performance. The support was not able to enhance the sensitivity for feature detection, a result which was also found in a change detection task by Williams and Simons (2000).

Rather, we suppose a detrimental effect of the support on the cognitive process of feature detection in interactive conditions: With support, the collaborators searched for features individually, while without support they searched collaboratively. Perhaps searching collaboratively helped to prevent illusions of simplicity (Nickerson, 1999) and also illusions of evidence that can arise if objects are physically co-present (Clark, 1996). However, in non-interactive conditions, the process data suggest a benefit of the support for feature detection.

The results further give rise to the supposition that the process of grounding might also be disturbed if the speaker individually searches for features and decides on names for them.
We implemented only one instructional support measure for both interactive and non-interactive conditions to allow for a conservative test of our measures. The support measures improved both the process of interactive and of non-interactive communication. Yet, descriptive tendencies suggest a different role of the support measures in the two conditions: In interactive conditions, dyads without support had no real disadvantage with regard to feature detection (percentage of identified features: mean\textsubscript{interactive, without} = 96.3; mean\textsubscript{interactive, with} = 95.7). In non-interactive conditions, speakers had no feedback or help from the addressee with regard to feature detection and support was quite helpful: Without support, they missed more features (percentage of identified features: mean\textsubscript{non-interactive, without} = 65.1; mean\textsubscript{non-interactive, with} = 77.4). In future research, the effect of more situation-specific support measures should be examined.

All in all, the developed support measures can be used as a starting point to help the worker and the mechanic introduced in our example to deal with problem-solving and communicative demands and to ensure an effective collaboration when repairing a complex engine via remote collaboration. Shared applications, as used in our first experiment, may be used to support the communicative demands: They can help to install referential identity, i.e. to make sure the workers are using the same words to refer to the same parts of the engine.

Instructional support measures, as used in our second experiment, may help to improve both the problem-solving as well as the communication process. However, our results suggest the need for inspecting carefully what types of subtasks are most successfully tackled individually and what subtasks have to be solved collaboratively.
REFERENCES


Supporting Remote Collaborative Problem-Solving

AUTHOR NOTE

Miriam Hansen, Department of Psychology, Goethe-University Frankfurt, Germany and Hans Spada, Department of Psychology, University of Freiburg, Germany.

This research was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) as part of the Virtual PhD Program (VGK) “Knowledge Acquisition and Knowledge Exchange with New Media”. We thank Richard Alterman, William Sid Horton, and Joe Krummenacher as well as two anonymous reviewers for helpful comments on earlier versions of this paper.

Correspondence concerning this article should be addressed to Miriam Hansen, Department of Psychology, Goethe-University Frankfurt, Postfach 11 19 32 D-60054 Frankfurt am Main, Germany. E-mail: hansen@paed.psych.uni-frankfurt.de
Table 1

Task characteristic and demands of concrete and abstract sets of pictures.

<table>
<thead>
<tr>
<th>Type of pictures</th>
<th>Concrete</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task characteristic</td>
<td>Familiar objects</td>
<td>Patterns/ geometrical figures</td>
</tr>
<tr>
<td>Demands</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual problem-solving</td>
<td></td>
<td>Search more error-prone</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Search less error-prone</td>
</tr>
<tr>
<td>Communicative</td>
<td></td>
<td>Easy to describe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Difficult to describe</td>
</tr>
</tbody>
</table>
Table 2

Means and standard deviations (in parentheses) of the process and performance measures in Experiment 1 for the factors type of pictures and mode of communication.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Type of pictures</th>
<th>Mode of communication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concrete</td>
<td>Abstract</td>
</tr>
<tr>
<td>Process measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified features (percentage)</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>“Feature not mentioned” errors</td>
<td>9.0 (7.3)</td>
<td>3.4 (3.1)</td>
</tr>
<tr>
<td>(frequency)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>“Name and position of feature” errors (frequency)</td>
<td>.6 (1.1)</td>
<td>2.4 (2.8)</td>
</tr>
<tr>
<td>“Frame of reference” errors</td>
<td>1.2 (3.1)</td>
<td>.2 (1.2)</td>
</tr>
<tr>
<td>(frequency)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Repeated picture descriptions</td>
<td>22.6 (6.1)</td>
<td>19.7 (3.6)</td>
</tr>
<tr>
<td>(frequency)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Warning signals/ clarification</td>
<td>5.2 (4.5)</td>
<td>7.4 (5.5)</td>
</tr>
<tr>
<td>questions asked (frequency)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Coordination activities</td>
<td>.6 (.9)</td>
<td>1.1 (1.1)</td>
</tr>
<tr>
<td>(frequency)</td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>
Table 2 (continued)

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Type of pictures</th>
<th>Mode of communication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concrete</td>
<td>Abstract</td>
</tr>
<tr>
<td>Performance measures</td>
<td>62.2 (26.7)</td>
<td>86.1 (14.4)</td>
</tr>
<tr>
<td>Pictures placed in correct position (percentage)</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Time needed to complete the task (in sec.)</td>
<td>600 (200)</td>
<td>972 (295)</td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>
Table 3

Means and standard deviations (in parentheses) of the process, time period, and performance measures in Experiment 2.

<table>
<thead>
<tr>
<th>Type of pictures</th>
<th>Amount of support</th>
<th>Mode of comm.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Concrete Level 1+2</td>
<td>Abstract Level 1</td>
</tr>
<tr>
<td>Process measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified features</td>
<td>78.8</td>
<td>90.4</td>
</tr>
<tr>
<td>(percentage)</td>
<td>(18.4)</td>
<td>(16.8)</td>
</tr>
<tr>
<td>“Feature not mentioned” errors</td>
<td>27.1</td>
<td>9.2</td>
</tr>
<tr>
<td>(frequency)</td>
<td>(14.3)</td>
<td>(12.6)</td>
</tr>
<tr>
<td>Features revealed before description</td>
<td>31.3</td>
<td>45.3</td>
</tr>
<tr>
<td>(percentage)</td>
<td>(32.1)</td>
<td>(46.0)</td>
</tr>
<tr>
<td>“Name and position of feature” errors (frequency)</td>
<td>0</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(1.8)</td>
</tr>
</tbody>
</table>

* * *
Table 3 (continued)

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Concrete</th>
<th>Abstract</th>
<th>Level 1+2</th>
<th>Level 1</th>
<th>No supp. Interactive</th>
<th>Non-interactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Frame of reference” errors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(frequency)</td>
<td>2.7</td>
<td>0</td>
<td>.5</td>
<td>1.4</td>
<td>2.1</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>(4.8)</td>
<td>(0)</td>
<td>(.8)</td>
<td>(3.2)</td>
<td>(2.4)</td>
<td>(2.3)</td>
</tr>
<tr>
<td></td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irrelevant features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(frequency)</td>
<td>13.7</td>
<td>3.1</td>
<td>6.9</td>
<td>8.3</td>
<td>9.9</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>(13.0)</td>
<td>(7.0)</td>
<td>(9.4)</td>
<td>(9.4)</td>
<td>(6.1)</td>
<td>(9.9)</td>
</tr>
<tr>
<td></td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complicated</td>
<td>.4</td>
<td>2.0</td>
<td>.6</td>
<td>1.2</td>
<td>1.8</td>
<td>.8</td>
</tr>
<tr>
<td>descriptions</td>
<td>(1.4)</td>
<td>(3.2)</td>
<td>(1.3)</td>
<td>(1.6)</td>
<td>(2.5)</td>
<td>(1.8)</td>
</tr>
<tr>
<td>(frequency)</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time period measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual time (in sec.)</td>
<td>358</td>
<td>550</td>
<td>705</td>
<td>411</td>
<td>155</td>
<td>347</td>
</tr>
<tr>
<td></td>
<td>(334)</td>
<td>(429)</td>
<td>(398)</td>
<td>(201)</td>
<td>(149)</td>
<td>(278)</td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(in sec.)</td>
<td>73</td>
<td>103</td>
<td>156</td>
<td>55</td>
<td>52</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>(89)</td>
<td>(115)</td>
<td>(102)</td>
<td>(72)</td>
<td>(96)</td>
<td>(104)</td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
Table 3 (continued)

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Concrete</th>
<th>Abstract</th>
<th>Level 1+2</th>
<th>Level 1</th>
<th>No supp.</th>
<th>Interactive</th>
<th>Non-interactive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictures placed in correct position (percentage)</td>
<td>32.8 (41.7)</td>
<td>72.2 (40.0)</td>
<td>51.7 (40.0)</td>
<td>57.2 (36.1)</td>
<td>49.4 (47.2)</td>
<td>80 (21.7)</td>
<td>25.6 (28.9)</td>
</tr>
<tr>
<td>Time needed for description and positioning (in sec.)</td>
<td>1064 (499)</td>
<td>1058 (472)</td>
<td>869 (371)</td>
<td>1103 (509)</td>
<td>1182 (538)</td>
<td>1136 (483)</td>
<td>986 (438)</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1. Examples of concrete (above) and abstract (below) pictures (those used in Experiment 2). We developed some of the pictures based on the children’s game Differix from Ravensburger Spieleverlag (Printed with permission).

Figure 2. Data of shared pictorial application from two dyads in Experiment 1, abstract task: one coded as used to support the individual problem-solving demands (right) and one as used to support the communicative demands (left) (Printed with permission of Ravensburger Spieleverlag).