

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

Manuscript - Pre-Print - Review Version

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. video-conferencing 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

1
2
3 notably difficult in remote settings due to the reduced or lack of physical copresence (Clark &
4 Brennan, 1991; Kraut, Fussell & Siegel, 2003). The worker and mechanic are only interacting
5
6 via an audio channel, and are neither able to see each other's facial expression nor use deictic
7
8 gestures for grounding (Clark & Krych, 2004; Fussell et al., 2004). Due to the various demands
9
10 on the collaborators, errors and problems are likely to occur and a satisfactory solution may not
11
12 be found within a short period of time. To fulfill the worker's desire for a good collaboration
13
14 outcome, their collaboration should be supported.

15
16
17
18
19
20 Such support may be realized by improving the technical environment (e.g. adding a
21
22 visual channel; Whittaker & O'Connell, 1997; providing shared workspaces; Tang, 1991;
23
24 Whittaker, Geelhoed & Robinson, 1993) or by structuring the interaction process (e.g.
25
26 collaboration scripts; O'Donnell & Dansereau, 1992; Baker & Lund, 1997). Collaboration scripts
27
28 have also been used to instruct collaboration partners on how to solve a remote collaborative task
29
30 (Rummel & Spada, 2005). With this instructional approach, a sustainable enhancement of the
31
32 collaborative process and results can be achieved. However, in order to design such instructional
33
34 support measures, we need to understand in detail what skills are needed to solve such a
35
36 collaborative task.
37
38
39

40 41 AIMS OF THE PAPER 42

43
44 With this paper, we aim to give a detailed picture of the relevant processes of
45
46 collaboration in our new remote picture-sorting task. We try to identify the individual problem-
47
48 solving and communicative demands of the task in order to develop effective instructional
49
50 support measures. Most research on collaboration and communication focusing on real-world
51
52 contexts does not take the findings of cognitive science sufficiently into account (see McNamara,
53
54 2006). A promising exception is the model recently presented by Horton and Gerrig (2005a),
55
56
57
58
59
60

1
2
3 which uses cognitive theories of memory to explain grounding and audience design processes.
4
5 Several studies have illustrated how this memory-based approach helps to improve the
6
7 understanding of communication processes (Horton, 2007; Horton & Gerrig, 2005b; Horton &
8
9 Spieler, 2007).
10
11

12
13 To build a foundation for instructional support, we also try to bridge the gap and bring
14
15 cognitive models together with models of communication. Therefore, the first part of this paper
16
17 presents a theoretical framework of the task demands, emphasizing the individual problem-
18
19 solving demands and the links to research on communication. In a first experiment, we examined
20
21 the impact of these theoretically derived demands on collaboration process and outcome.
22
23

24
25 Our second aim is to evaluate support measures for the remote collaborative picture-
26
27 sorting task. In the first experiment, we tested simple *technical support measures* (shared
28
29 applications), while in the second experiment, we evaluated *instructional support measures*
30
31 (model collaboration and collaboration script) including support of the individual problem-
32
33 solving and the communicative demands.
34
35

36 THE COLLABORATIVE PICTURE SORTING TASK

37
38 Studies of collaboration mostly use realistic and therefore quite complex settings (e.g.
39
40 Fussell et al., 2000; Olson & Olson, 2000; Rummel & Spada, 2005) and do not allow the
41
42 underlying processes to be examined in depth. Therefore, we used a more restricted remote
43
44 collaborative task that enables a more detailed analysis. In the following, we briefly describe the
45
46 task before presenting the framework of demands. In the following, we briefly describe the
47
48 task before presenting the framework of demands.
49

50
51 In the broadest sense, the task used in our studies can be classified into a “general class of
52
53 ‘mentoring’ collaborative physical tasks” (Kraut et al., 2003; p. 16), in which one person
54
55 manipulates pictures and receives instructions from another person about where to place them.
56
57
58
59
60

1
2
3 The task accomplished by the worker and mechanic in the example described above also fits into
4 this category. The structure of these tasks is similar to the referential communication task
5
6 (Krauss & Weinheimer, 1966), which has often been used to study communication (e.g. Brennan
7
8 & Clark, 1996; Horton & Keysar, 1996). In a referential communication task, a speaker typically
9
10 describes some clearly different pictures to an addressee. In previous referential communication
11
12 studies, impacts of task characteristics as codability, discriminability, and familiarity of the
13
14 material (Bortfeld, Leon, Bloom, Schober & Brennan, 2001; Gergle, Kraut & Fussell, 2006;
15
16 Horton & Gerrig, 2002; Hupet, Seron & Chantraine, 1991) have been found to have an influence
17
18 on the communicative effort needed to achieve mutual understanding.
19
20
21
22
23

24
25 However, our collaborative picture-sorting task differed on one important point from the
26
27 classical task: Our participants needed to conduct a visual search to detect relevant differences
28
29 between quite similar pictures. Thereby, individual problem-solving demands (visual search
30
31 processes) were added to the communicative demands: As in realistic collaborative physical
32
33 tasks, this corresponded to a dual task situation in which the participants had to accomplish
34
35 individual problem-solving activities *in parallel* to communication. In contrast to previous work,
36
37 our task allowed to scrutinize the impact of task characteristics on both individual problem-
38
39 solving and communicative demands.
40
41
42

43
44 In both experiments included in this paper, two persons sitting in two different rooms
45
46 jointly solved a picture-sorting task. One of the participants assumed the role of *speaker* and the
47
48 other took the role of *addressee* (synonymous terms are *director* and *matcher*). In the following,
49
50 we adopt the convention of assuming a female speaker and a male addressee. The task was
51
52 presented on two displays and oral communication between speaker and addressee was possible
53
54 via an audio link.
55
56
57
58
59
60

1
2
3 On the speaker's display, a number of pictures were presented that differed only in terms
4 of minor details. The speaker had to detect the relevant features and describe them and their
5 order to the addressee. The addressee saw the pictures in a random order and had to arrange them
6 according to the speaker's description. The addressee could rearrange the pictures on the target
7 area by using the mouse (drag and drop). Because the differences between the pictures were very
8 small, the participants first had to search for them. This component of feature detection
9 accomplished in parallel to communication constitutes the main difference from the classical
10 referential communication task, in which the task demand consists in the verbal description of
11 clearly different pictures or objects. This additional individual problem-solving demand makes
12 the task more comparable to realistic collaborative tasks in which communication often has to
13 take place in parallel to individual problem-solving processes (e.g. the worker communicating on
14 the phone with the mechanic while searching for a broken part, or an air traffic controller
15 monitoring different displays while interacting with pilots). Furthermore, the task is comparable
16 to collaborative settings with distributed resources or distributed skills, such as expert-layperson
17 communication (e.g. Bromme, Jucks, & Wagner, 2005; Nückles & Ertelt, 2006) or hidden-
18 profile tasks (e.g. Stasser & Titus, 1985), because the speaker has information about the required
19 target order of the pictures that the addressee does not have.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

43 FRAMEWORK OF INDIVIDUAL PROBLEM-SOLVING DEMANDS AND 44 COMMUNICATIVE DEMANDS (IproCo) 45 46 47

48 The theoretical framework we propose aims to bring together cognitive models as well as
49 assumptions from communication pragmatics. It highlights the parallelism of individual
50 problem-solving demands and communicative demands. IproCo includes *task characteristics*
51 (concrete vs. abstract pictures) and *situation characteristics* (e.g. interactive vs. non-interactive
52
53
54
55
56
57
58
59
60

1
2
3 communication) as important factors influencing the two demands. Further, it stresses the
4
5 *awareness* of the task and situation-specific difficulties as a crucial prerequisite for a successful
6
7 collaboration.
8
9

10 Although the framework is tailored to the collaborative picture-sorting task that we used,
11
12 it can easily be applied to other collaborative tasks.
13
14

15 Individual Problem-Solving Demands

16
17 In our task, the individual problem-solving demands consist in visual search processes: In
18
19 order to identify the relevant features (e.g. “position of the cat”) and feature values (e.g.
20
21 “position of the cat: left vs. right”) that differ between the pictures, a serial visual search (e.g.
22
23 Treisman & Gelade, 1980) has to occur. This requires the allocation of attention and is a time-
24
25 consuming and error-prone process because focused attention is only a limited resource and
26
27 cannot be directed to all details at the same time (Treisman & Schmidt, 1982). When observing a
28
29 scene or looking at pictures, changes in objects or parts of objects are therefore often not
30
31 detected – an effect called *change blindness* (e.g. Rensink, O'Regan, & Clark, 1997). This effect
32
33 was found independently of the occurrence of saccades if attention was not allocated to the
34
35 changing part of the picture (O'Regan, Deubel, Clark, & Rensink, 2000; Rensink et al., 1997).
36
37 Such an effect was shown not only for change detection tasks, but also if the participants
38
39 searched for differences in simultaneously presented pictures (Brunel & Ninio, 1997; Hansen,
40
41 Spada, & Schneider, 2007; Scott-Brown, Baker, & Orbach, 2000). Change blindness occurs also
42
43 in social settings (Simons & Levin, 1998) and has even been identified as a possible reason for
44
45 inaccuracy in witnesses to crime (Davis, Loftus, Vanous, & Cucciare, 2007).
46
47
48
49
50
51
52

53 The visual salience of the differences, such as size, location, or color, was found to
54
55 influence the detection of change (Smilek, Eastwood, & Merikle, 2000; Williams & Simons,
56
57
58
59
60

1
2
3 2000). However, differences in low-level factors such as visual salience are not the only
4
5 determinants for the allocation of attention. Rather, high-level factors such as *familiarity* or
6
7 semantic differences have been shown to highly influence the detection of changes (O'Regan et
8
9 al., 2000; Pearson & Schaefer, 2005; Rensink et al., 1997). Rensink and colleagues (Rensink et
10
11 al., 1997; O'Regan et al., 2000) used the *degree of interest* of a picture part as an
12
13 operationalization and found that changes are more likely to be detected if they occur on parts of
14
15 central interest than of marginal interest.
16
17
18

19
20 In both experiments reported here, we used two types of pictures that differed according
21
22 to their *task characteristics*, i.e. familiarity.
23
24

25 Insert Figure 1 about here
26

27 The concrete pictures contained familiar or known objects, whereas the abstract ones
28
29 depicted geometrical figures or patterns of figures (see Figure 1). The search for features in a
30
31 concrete picture resembles the process of finding differences in a so called 'find the difference'
32
33 task, whereas searching for features in abstract tasks is more comparable to processes of pattern
34
35 recognition. Table 1 lists the task characteristics and demands that were assumed to vary
36
37 between concrete and abstract pictures.
38
39

40
41 Insert Table 1 about here
42
43

44 To verify the differences in the task characteristics, we presented the pictures to 46
45
46 students and asked them to rate the pictures on a 5-point scale regarding their familiarity (see
47
48 Snodgrass & Vanderwart, 1980). The participants rated the concrete pictures ($M = 4.1$; $SD = .96$)
49
50 as significantly more familiar than the abstract pictures ($M = 2.1$; $SD = .90$; $t[45] = 11.35$, $p <$
51
52 $.001$).
53
54
55
56
57
58
59
60

Supporting Remote Collaborative Problem-Solving 10

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 permit more than one meaning. She has to learn through the reactions of the addressee *when*
4
5 audience design is necessary (Horton & Gerrig, 2002) and then has to be *aware* of such task and
6
7 situation characteristics.
8
9

10 One task characteristic determining the communicative demands is – again – the
11
12 *familiarity* of an object: Familiar or concrete objects with a high degree of linguistic codability
13
14 are easier to describe (e.g. Barry, Morrison & Ellis, 1997; Bortfeld et al., 2001; Hupet et al.,
15
16 1991; Oldfield & Wingfield, 1965) and the range of possible expressions is much smaller than
17
18 for unknown or abstract objects (Bertholet & Spada, 2005; Horton & Gerrig, 2002). With
19
20 unfamiliar objects, the speaker has to invent terms and to introduce them to the addressee in
21
22 order to establish referential identity (Isaacs & Clark, 1987); the collaborators have more
23
24 difficulties in negotiating a conceptual pact (Brennan & Clark, 1996). Therefore, we expect
25
26 abstract pictures to be more difficult to describe (see Table 1).
27
28
29
30
31

32 One situation characteristic determining the communicative demands is the *mode of*
33
34 *communication*. While interactive communication is the main setting for language use (e.g.
35
36 Pickering & Garrod, 2004), there are also settings involving less interactive communication or
37
38 settings without feedback options such as lectures and speeches or written communication (we
39
40 will use the term *non-interactive* to refer to less-interactive settings). Non-interactive settings do
41
42 not provide the speakers with the opportunities to learn about the need for audience design, as
43
44 the speaker does not receive feedback and cannot be certain of having been understood by the
45
46 addressee. Grounding (Clark & Brennan, 1991; Clark & Schaefer, 1987) is not possible and
47
48 more communicative effort is needed to overcome the hindering influences (Bromme, Hesse, &
49
50 Spada, 2005; Münzer & Borg, 2007). Non-interactive settings also affect the addressee's
51
52
53
54
55
56
57
58
59
60

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

1
2
3 randomly grouped into dyads and assigned the role of speaker or addressee. Participants did not
4
5 know each other prior to the study.
6
7

8 *Design and procedure*

9
10 A 2x3 factor design was implemented, with the mode of communication (interactive vs.
11 restricted communication) and the availability of a shared application (pictorial, textual, or none)
12 as between-subject factors and with the type of pictures (concrete vs. abstract) as an additional
13 within-subject factor.
14
15
16
17

18
19 Each dyad was assigned randomly to one of the six conditions. Each participant received
20 instructions in the form of a manual explaining the use of the technical environment and the
21 procedure. The dyad then performed a training task to familiarize themselves with the technical
22 environment. They were instructed to complete the tasks as accurately and quickly as possible.
23
24 During the experimental phase, each dyad was required to solve two concrete and two abstract
25 tasks. All dyads received the tasks in the same sequence. The speaker described the pictures and
26 their positions and the addressee placed the pictures on his or her display accordingly. After
27 completing each task, the participants received feedback about the number of pictures they had
28 placed in the correct position and were informed about the features that had been relevant in the
29 task (see next section).
30
31
32
33
34
35
36
37
38
39
40
41
42

43 *Type of pictures.* The concrete tasks contained familiar objects, while the abstract tasks
44 contained patterns or geometrical figures (see Table 1). The number of relevant features differed
45 depending on the task (three and four features in concrete, four and five in abstract tasks);
46 whereas the relevance of a feature was determined by the necessity to describe its value in order
47 to identify a picture unambiguously. For one of the concrete tasks - the cat-task (see Figure 1) -
48 these features were: the position of the cat, the position of the cat's head, the direction of the
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 cat's whiskers, and the form of the mouse's tail (mouse on the bottom). An irrelevant feature
4
5 was, for example, the position of the cat's tail in addition to the position of the cat, as these
6
7 features varied conjointly. However, if a speaker did choose to describe the position of the cat's
8
9 tail *instead* of describing the position of the cat, this was considered as a relevant feature.
10
11

12
13 *Mode of communication.* Depending on the condition, the addressee was either able to
14
15 talk or not. In the latter case, the addressee did, however, have the possibility to send an audio
16
17 warning signal to the speaker if there was any trouble in understanding. This enabled a
18
19 comparison of interactive and restricted communication settings.
20
21

22
23 *Shared application.* Additionally, a pictorial or textual shared application was each
24
25 presented in one interactive and one restricted communication condition. These applications
26
27 influencing the situation characteristics were introduced as a second factor because their
28
29 availability may lead to a kind of copresence (Clark & Brennan, 1991) and thus facilitate
30
31 communication. The pictorial shared applications contained a picture of the respective task by
32
33 way of example and had a pointing and drawing feature that enabled deictic gestures. The textual
34
35 shared applications consisted of a shared text editor and allowed the relevant differences to be
36
37 noted down. These notes were always present on both screens and could be looked at again
38
39 during the collaboration process. In the conditions with a shared application, only the speaker
40
41 was able to make drawings or write text, but these were also visible to the addressee. However,
42
43 the speakers were not obliged to use the shared application to support their explanations, and
44
45 were merely informed about the additional possibility of using it.
46
47
48
49

50 *Measures*

51
52
53 Three sets of data were collected to examine the process and the outcome of
54
55 collaboration: (1) audio recordings of the verbal communication (process measures), (2)
56
57
58
59
60

1
2
3 performance data (performance measures), and (3) data of the shared applications (drawings or
4
5 text).
6
7

8 *Process measures.* To examine whether the postulated demands affected the
9
10 communication process, a coding scheme was developed consisting of categories for the
11
12 individual problem-solving demands and the communicative demands.
13
14

15 The first group included the *percentage of identified features*. This category was a rate
16
17 measure, as there was a different amount of relevant features depending on the task (three and
18
19 four features in concrete, four and five in abstract tasks). All other categories were frequency
20
21 measures, because they related to single picture descriptions (i.e. a set of the speaker's utterances
22
23 describing one picture). The second category for the individual problem-solving demand – *the*
24
25 *number of "feature not mentioned" errors* – counted the number of picture descriptions in which
26
27 relevant features were not mentioned. Note that the number of picture descriptions could be
28
29 higher than nine, in case the speaker described some pictures more than once.
30
31
32
33

34 During coding, we made one important change to the categories for the communicative
35
36 demands: Initially, we counted all ambiguous descriptions reflecting failures to establish
37
38 referential identity. Then, we realized that this communicative error category was too broad and
39
40 in fact comprised two distinct types of errors: ambiguous descriptions regarding the feature's
41
42 name or position and ambiguous descriptions regarding the frame of reference. A spatial
43
44 description can either be viewer-centered, object-centered, or environment-centered (see Levelt,
45
46 2003). Notably, in descriptions of concrete pictures, ambiguities may be expected regarding the
47
48 frame of reference used. A viewer-centered and an object-centered frame of reference can lead to
49
50 two opposite meanings (e.g. "the cat is looking to the left" – described from the cat's or the
51
52 viewer's perspective?). Therefore, we included two distinct categories, both reflecting failures to
53
54
55
56
57
58
59
60

1
2
3 establish referential identity: *number of “name and position of feature” errors* and *number of*
4
5 *“frame of reference” errors*. Further, we included the *number of repeated picture descriptions*,
6
7 which counted a set of the speaker’s utterances describing one picture. In restricted
8
9 communication conditions, we counted the *number of warning signals* and in interactive
10
11 conditions, we counted the *number of clarification questions* the addressee asked. We also
12
13 counted the *number of coordination activities*, i.e. the number of times the speaker used meta-
14
15 language to talk about the way in which to describe the features.
16
17
18

19
20 It should be noted that some categories are not independent of each other: Of course, if a
21
22 speaker did not describe a relevant feature (individual problem-solving demand), then she could
23
24 not commit errors in the description (communicative demand). The audio recordings of the
25
26 verbal communication were analyzed using these categories. Ten percent of the verbal data was
27
28 coded by a second rater to enable the calculation of inter-rater reliability. The consistency of the
29
30 coding was medium to high, with mean intra-class correlation (e.g. Shrout & Fleiss, 1979) of .81,
31
32 indicating that the coding scheme could easily be used.
33
34
35

36
37 *Performance measures.* The quality of the joint solution was measured by the *number of*
38
39 *pictures placed in the correct position* and the *time needed to complete the task*.
40

41
42 *Data of shared applications.* In terms of the data gathered from the shared applications,
43
44 the number of dyads that made use of them was counted. This data was then either coded as
45
46 supporting the individual problem-solving demands (feature identification/ remembering the
47
48 features), or as supporting the communicative demands (describing the feature values of each
49
50 picture). In the first case, the speaker had marked or written down all features at the beginning of
51
52 a task in order to make it clear what *features* generally differed between the pictures (e.g. “the
53
54 mouse’s tail, the position of the cat...”). Whereas in the second case, the speaker had used the
55
56
57
58
59
60

1
2
3 shared application to mark or write down the *feature values* of each picture (e.g. “In picture 1,
4 mouse’s tail in S-shape, cat on the left...”) during the description of the nine pictures.
5
6
7

8 Results and Discussion

9 *Process and performance data*

10
11
12 A multivariate analysis of variance (MANOVA) with repeated measures (for the factor
13 type of pictures) was conducted to test the influence of the three factors (type of pictures, mode
14 of communication, shared application) on the process and performance measures. There was an
15 effect of the type of pictures ($F[9, 46] = 20.6, p < .01, \eta^2 = .80$), an effect of the communication
16 mode ($F[9, 46] = 12.2, p < .01, \eta^2 = .71$), but no effect of the shared application. The lack of
17 effect of the shared application may be due to the fact that not all dyads of the respective
18 conditions used them, as their usage was not obligatory. Eighty percent of the dyads in the
19 respective conditions used it for at least one task, but the number of dyads using the pictorial
20 shared application differed notably from task to task and from condition to condition (varying
21 between 10 and 90 percent). We will present the rate of usage in more detail in the next section
22 (data of the shared application).
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 The MANOVA revealed also a significant interaction between type of pictures and mode
39 of communication ($F[9, 46] = 2.9, p < .01, \eta^2 = .36$).
40
41
42

43 Table 2 contains means and standard deviations of all variables for the factors type of
44 pictures and communication modes. As it had no significant effect, the factor shared application
45 is not included in Table 2. In the following, we will describe the main effects for the factors type
46 of pictures (second and third columns of Table 2) and mode of communication (last two
47 columns), as well as for the significant interaction (not included in Table 2).
48
49
50
51
52
53
54

55 Insert Table 2 about here
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

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.

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

Supporting Remote Collaborative Problem-Solving 20

1
2
3 problem-solving demands of feature detection were indeed more important in concrete tasks,
4
5 leading to fewer correctly placed pictures. In abstract tasks, the speakers had to make additional
6
7 effort to engage in audience design; as a consequence, more time was spent in completing the
8
9 tasks. This can be interpreted as an illusion of simplicity effect (Nickerson, 1999) for concrete
10
11 pictures: The familiarity of the concrete pictures led to the deceptive impression that the
12
13 differences were easy to detect and the pictures easy to describe. Therefore, the participants took
14
15 less time for search and description and committed more errors.
16
17
18
19

20 To scrutinize this effect, we asked 46 students about their impressions of each picture
21
22 from Experiment 1: They had to rate on a 5-point scale their impression of how easy it would be
23
24 to find the differences and how easy to describe them. The participants attributed significantly
25
26 fewer difficulties to finding the features for the concrete ($M = 2.0$; $SD = .79$) than for the abstract
27
28 pictures ($M = 3.9$; $SD = .79$; $F[1] = 140.1$, $p < .01$, $\eta^2 = .76$). Comparable results were found for
29
30 describing the features (concrete: $M = 1.9$; $SD = .75$; abstract: $M = 3.7$; $SD = .87$; $F[1] = 217.9$, p
31
32 $< .01$, $\eta^2 = .83$).
33
34
35

36 *Main effects of “mode of communication”.* Starting again with the process measures,
37
38 there were more features identified ($F[1, 54] = 5.2$, $p < .05$, $\eta^2 = .09$), fewer “name and position
39
40 of feature” errors made ($F[1, 54] = 3.1$, $p < .05$, $\eta^2 = .06$), more questions asked ($F[1, 54] = 46.5$,
41
42 $p < .01$, $\eta^2 = .46$), and more coordination activities carried out ($F[1, 54] = 9.9$, $p < .01$, $\eta^2 = .16$) in
43
44 interactive conditions.
45
46
47

48 With respect to the performance data, the dyads in interactive conditions placed
49
50 approximately one picture more in the correct position (ten percent points more than in abstract
51
52 tasks; $F[1, 54] = 3.1$, $p < .05$, $\eta^2 = .11$) and took roughly 100 seconds less to complete the task
53
54 ($F[1, 54] = 6.8$, $p < .01$, $\eta^2 = .06$). This is in line with previous findings showing the additional
55
56
57
58
59
60

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

1
2
3 data was either coded as *used to support the individual problem-solving demands*, if the speaker
4 used the shared application to support feature identification (What are the *features* that differ
5 between the pictures?), or as *used to support the communicative demands*, if the speaker used the
6 shared application to describe the feature value of each single picture (What are the *feature*
7 *values* of picture 1, picture 2, ...?). Figure 2 exemplarily shows data of the shared pictorial
8 application for both types of usage.
9
10
11
12
13
14
15
16

17
18 Insert Figure 2 about here
19

20 To further illustrate the usage of the shared application, consider the following
21 descriptions, each corresponding to one of the two pictures in Figure 2. The speakers drew into
22 the shared application while pronouncing the underlined words.
23
24
25
26

27 (1) Coded as *used to support the individual problem-solving demands* (What are the
28 *features* that differ between the pictures?):
29
30

31 S: You have to consider the red fields, the broken and the continuous lines. First picture:
32 Outer red field between nine and eleven. [...]
33
34
35

36 (2) Coded as *used to support the communicative demands* (What are the *feature values* of
37 picture 1, picture 2, ...?):
38
39

40 S: I'm starting with the first one. Well, there is a circle and this line here, only this line is
41 black... and then this line...this... this... and this. And this circle thing here. And what's
42 red, it is this box here, this, and this – they are red.
43
44
45
46
47

48 In both examples, the pictorial shared application was used to establish referential
49 identity with less effort. This facilitation becomes clear when comparing the above examples
50 with the following description made by a speaker who had no shared application available:
51
52
53
54
55
56
57
58
59
60

Supporting Remote Collaborative Problem-Solving 24

1
2
3 (3) S: First picture: On the right, outer part, there is this type of red quarter. Er – no – I
4
5 mean on the left outer part. Excuse me. Not really a quarter but, anyway – the red thing in
6
7 the outer circle is on the left side.
8
9

10 A: Okay. Bottom left?

11
12 S: No, upper part. More – er – half way in between.
13
14

15 Regarding the type of use, both shared applications were mostly used to support the
16
17 communicative demands, i.e. the speaker used the shared application to describe the feature
18
19 value of each single picture instead of using it to support feature identification. We will discuss
20
21 practical implications of these findings for the supply of shared applications in the overall
22
23 discussion.
24
25

26
27 *Implications for designing instructional support measures*
28

29 As one of our aims was to develop effective instructional support measures based on the
30
31 findings of first experiment, what did we learn that can help in the development of these
32
33 measures? First, the results support the theoretical assumption of the IproCo framework:
34
35 Individual problem-solving and communicative demands both have an impact on the
36
37 collaboration process and outcome. As predicted, they are influenced by the familiarity of the
38
39 task characteristics, with higher individual problem-solving demands arising in concrete tasks
40
41 and higher communicative demands arising in abstract tasks. They are further influenced by the
42
43 situation characteristics mode of communication (interactive/ restricted), with higher demands in
44
45 restricted communication conditions. Moreover, the process analysis helped to identify a second
46
47 task characteristic: the possibility for different frames of reference. Thus, a sustainable support
48
49 should aim at reducing the number of errors related to both demands by improving the
50
51 awareness: For concrete pictures, it should be emphasized to the participants that they should
52
53
54
55
56
57
58
59
60

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

Supporting Remote Collaborative Problem-Solving 26

1
2
3 they had just learned was provided during the collaboration. This combination should promote
4
5 effective collaboration, as we expect the model collaboration video to have positive effects on
6
7 the participants' attention and motivation (Bandura, 1977) and the collaboration script to
8
9 enhance the memory for the relevant information.
10
11

12 The instructional measures (model instruction and script) were aimed at augmenting the
13 awareness for the task characteristics necessitating audience design and to transform the parallel,
14
15 dual task structure into a sequential one: (1) searching for features, (2) finding feature's names as
16
17 to describe the relevant features in an understandable way, and (3) describing the single pictures.
18
19 The measures contained two levels of support: level 1, corresponding to the subtask "searching
20
21 for features" and thereby supporting the individual problem-solving demands, and level 2,
22
23 corresponding to the subtask "finding names for the features" and thus supporting the
24
25 communicative demands. Each support level contained hints for dealing with the respective
26
27 demand and, moreover, each level introduced one subtask to force the speakers to follow the
28
29 sequential task structure: marking the features in an individual picture editor (level 1) and writing
30
31 the features' names into an individual text editor (level 2). It should be noted, that these editors
32
33 were *no* shared applications as in Experiment 1, but *individual* editors. In contrast to Experiment
34
35 1, they were not introduced to be used as additional communication channel but to guarantee the
36
37 intended sequential structure of the task with one demand after another to be dealt with: Marking
38
39 the features in the individual picture editor should help to identify all relevant features and
40
41 values. Writing down the features' names should help to formulate appropriate descriptions and
42
43 – combined with hints for the description – to be sensitized for pitfalls hindering audience
44
45 design.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 The results of the first experiment showed restricted communication to be a very difficult
4 undertaking. In order to test for effects of support, we included interactive as well as non-
5 interactive conditions in Experiment 2 and provided both with instructional support measures.
6
7
8 This time, we established a more rigorous non-interactive condition: Speaker and addressee in
9 non-interactive conditions performed the tasks not at the same time, but rather one after another.
10
11 The speaker recorded his or her description and the audio recording was given to the addressee
12 later on. However, we provided both interactivity conditions with the same support measures, as
13 we focused on the awareness for task characteristics.
14
15
16
17
18
19
20
21

22 The second experiment aimed at answering the following research questions:

23
24 (1) Do the instructional support measures improve collaboration process and outcome?
25
26

27 We expected to reveal an impact of the instructional support measures in both process and
28 outcome, with complete support (level 1 and 2) being most beneficial.
29
30

31
32 (2) Is there a difference in the impact of the measures in non-interactive compared to
33 interactive communication conditions? We expected our support measures to be helpful for both
34 interactive and non-interactive conditions.
35
36
37

38 Method

39 *Participants*

40
41
42
43 Ninety-six students (48 dyads) of the University of Freiburg, Germany participated in the
44 study. Students of psychology were excluded. Thirty-six of the participants were male and 60
45 were female. The participants had an average age of 24.15 years ($SD = 4.4$, range = 18 to 48) and
46 all were German native speakers. The experiment lasted for 90 to 120 minutes and each
47 participant received 15 Euros for taking part. The participants were randomly grouped into
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 dyads, assigned the role of speaker and addressee, and randomly assigned to one of the six
4
5 conditions. Participants did not know each other prior to the study.
6
7

8 *Design and procedure*

9
10 A 2x3 factor design was used, with the amount of support (no support/ level 1 support/
11 level 1 and 2 support) and the mode of communication (interactive/ non-interactive) as between-
12 subject factors. Again, the type of pictures (concrete/ abstract) was varied as an additional
13 within-subject factor.
14
15
16
17
18

19
20 Prior to the collaboration, the participants individually received instructions including
21 technical advice in the form of an on-screen video. After watching the on-screen video, the dyads
22 performed a training task in order to familiarize themselves with the technical environment and
23 the subtasks. During the experimental phase, each dyad was required to solve two concrete and
24 two abstract tasks but in different sequences: To control for sequence effects, four different task
25 sequences were given. However, as an ANOVA revealed no effect of the task order, this factor
26 will not be taken into account in the subsequent analyses. As in the first experiment, 9 pictures
27 had to be described and positioned. Yet, this time, the addressee had 16 pictures per task to
28 choose from on his or her display. The speaker still had 9 target pictures but saw 7 additional
29 pictures on his or her display. As more pictures were presented, the difficulty of choosing the
30 correct picture from the remaining ones was therefore still present for the 8th and 9th pictures.
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45
46 *Amount of support.* Depending on the condition, the on-screen video included only
47 technical advice (no support conditions) or additionally contained the support of the problem-
48 solving demands (level 1 support) or of both demands (level 1 and 2 support). Level 1 support
49 contained hints for feature search (e.g. pointing out the counter-intuitive higher difficulty of
50 feature detection in concrete task) and introduced the first individual subtask (marking the
51
52
53
54
55
56
57
58
59
60

1
2
3 differences in the individual picture editor). Moreover, level 1 and 2 support included hints for
4
5 dealing with the communicative demands (e.g. demonstrating the utility of explicit grounding of
6
7 the features and of specifying the frame of reference used) presented as model collaboration, by
8
9 displaying textboxes with the model speaker's descriptions. Additionally, the second individual
10
11 subtask was introduced (writing the features' names in the individual text editor). By making the
12
13 division into individual and collaborative subtasks explicit, the instructional support measures
14
15 were meant to overcome the collaborators' tendency to solely engage in joint activities and
16
17 forget about individual work phases (Hermann, Rummel & Spada, 2001). The parallelism of the
18
19 two demands should further be fostered by the support measures.
20
21
22
23

24
25 *Mode of communication.* In the non-interactive conditions, the speaker and addressee
26
27 performed the tasks not at the same time, but rather one after another. The speaker had a
28
29 recording device on the display and was able to start and stop recording the explanations to the
30
31 addressee as desired. A microphone was positioned on the table next to the monitor. The
32
33 description of each speaker was randomly assigned to one addressee, who later arranged the
34
35 pictures according to the recorded descriptions. The addressee had an audio-player device on the
36
37 display and was able to start, stop and rewind the recording of the speaker's descriptions.
38
39 However, speakers and addressees in both the non-interactive and interactive conditions were
40
41 told to proceed as accurately and quickly as possible.
42
43
44

45 46 *Measures*

47
48 Two sets of data were collected to examine the process and the outcome of collaboration:
49
50 (1) audio recordings of the verbal communication (process measures, time periods) and (2)
51
52 performance data (performance measures).
53
54
55
56
57
58
59
60

Supporting Remote Collaborative Problem-Solving 30

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

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.

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 understanding and corresponds to a part of the grounding process. According to Clark and
4
5 Schaefer (1987), the grounding process consists of a presentation and an acceptance phase. In
6
7 asynchronous conditions, the information phase corresponded to the presentation phase, whereas
8
9 in synchronous conditions, it also included the acceptance phase, during which the addressee
10
11 gives evidence of understanding or misunderstanding.
12
13

14
15 In conditions without support, there was no explicit direction for structuring the process
16
17 into these time periods. And in level 1 support conditions, only the individual period was
18
19 instructed. However, we discerned the individual time before the speaker started the description
20
21 and the information time for all dyads, as we believed them to be necessary steps of the problem-
22
23 solving and communication process and some speakers in no or level 1 support conditions made
24
25 spontaneously use of them. Speakers that did not use an individual time period performed feature
26
27 search and description processes simultaneously.
28
29
30

31
32 *Performance measures.* As in the first experiment, the performance measures included
33
34 the number of pictures placed in the correct position and the time needed for picture description
35
36 and positioning. Thus, the total time on task was composed of the individual time, the
37
38 information time, and the time needed for picture description and positioning.
39
40

41 *Results and Discussion*

42
43 We computed a MANOVA with repeated measures (for the factor type of pictures) in
44
45 order to test the influence of the three factors on process and performance measures. There was
46
47 an effect of the type of pictures ($F[11, 32] = 12.8, p < .01, \eta^2 = .81$), an effect of the amount of
48
49 support ($F[22, 66] = 2.8, p < .01, \eta^2 = .48$) and an effect of the mode of communication ($F[11,$
50
51 $32] = 9.5, p < .01, \eta^2 = .77$). There were no significant interactions. Table 3 contains means and
52
53 standard deviations for all variables. In the following, we will describe the main effects for the
54
55
56
57
58
59
60

Supporting Remote Collaborative Problem-Solving 32

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).

Insert Table 3 about here

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

1
2
3 problems occurred mostly in conditions without support: The speakers searched for the features
4
5 and started with the description in parallel, leading to more errors.
6
7

8 With regard to the time periods, the dyads with complete support separated feature search
9
10 (individual time), grounding of features and values (information time), and description of single
11
12 pictures (time needed for picture description and positioning), as intended by the support
13
14 measures: These dyads took the highest amount of individual time, followed by dyads with
15
16 support of the individual problem-solving demands and finally the dyads without support ($F[2,$
17
18 $42] = 41.1, p < .01, \eta^2 = .66$). The information time of dyads with complete support was three
19
20 times longer than of dyads in the other two conditions ($F[2, 42] = 8.7, p < .01, \eta^2 = .29$). In this
21
22 information period, the speakers listed the features they had found, presented the names for
23
24 features and values, and defined the frame of reference for their descriptions. In conditions
25
26 without support or only with level 1 support, the lack of explicit information brought problems
27
28 and irritations. The following dialog of a dyad without support illustrates this (concrete task
29
30 “cat”, first picture; see Figure 2).
31
32
33
34
35

36 (4) S: The cat's head is directed upwards to the left.

37 A: Head tilted to the left.

38 S: No, wait. Not tilted to the left! Directed upwards to the left.

39 A: But that's tilted to the left.
40
41
42
43
44
45

46 The performance measures, however, did not reflect the impact of the measures
47
48 unanimously: Dyads without support took more time for description and positioning of the
49
50 pictures than dyads with support of the individual demands as well as dyads with complete
51
52 support. Dyads with complete support were faster than dyads in the other conditions ($F[2, 42] =$
53
54
55
56
57
58
59
60

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

1
2
3 familiarity and possibility for different frames of reference (task characteristics). Consequently,
4 they should also vary in the amount of individual problem-solving and communicative demands.
5
6 The results of both experiments confirmed the expected effects on process and outcome of the
7
8 collaboration: In concrete tasks, the feature search (individual problem-solving demands) was
9
10 more error-prone, and more “frame of reference” errors (one aspect influencing communicative
11
12 demands) occurred. In abstract tasks, the communicative demands proved to be predominant in
13
14 the collaboration process. The outcome of the collaboration may seem to be a time-accuracy
15
16 trade-off, as there were fewer pictures placed in the correct position in concrete tasks and also
17
18 less time used. Indeed, we interpret this in the light of the different demands: As expected, the
19
20 features were more difficult to find in concrete tasks, more features were overlooked, and fewer
21
22 pictures were placed correctly. In abstract tasks, the speakers spent more time for the
23
24 descriptions due to the additional effort for audience design (Horton & Gerrig, 2002). We
25
26 interpreted this as an illusion of simplicity (Nickerson, 1999) that occurred for concrete pictures:
27
28 The familiarity of the concrete pictures led to the deceptive impression that the differences were
29
30 easy to detect and the pictures easy to describe. Therefore, the participants took less time for
31
32 search and description and committed more errors. Interactivity is one example of situation
33
34 characteristics, which were also expected to influence individual problem-solving and
35
36 communicative processes. In both experiments, we compared conditions enabling normal
37
38 bidirectional, interactive communication to conditions enabling no or only restricted feedback
39
40 from the addressee, i.e. restricted communication or non-interactive conditions. This allowed an
41
42 examination of the individual achievement of the speaker without interaction and feedback
43
44 effects in non-interactive or restricted conditions. Furthermore, the collaboration and problem-
45
46 solving processes could be compared in order to examine whether speakers in non-interactive
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 conditions could compensate for the lack of feedback. Not surprisingly, dyads in interactive
4
5 conditions had better results than dyads in restricted or non-interactive conditions. As the
6
7 analysis of the process data showed, this was mostly due to the fact that, working together, two
8
9 people identified more features. Furthermore, they had the possibility to speak about problems
10
11 and the planned procedure. The illusion of simplicity discussed above was even more important
12
13 in non-interactive conditions, since the number of features that the speaker did not identify at all
14
15 was especially high in concrete tasks. To summarize, less interactive or non-interactive settings
16
17 amplified the predominant demands – individual problem-solving in concrete and
18
19 communicative demands in abstract tasks (see Table 1).
20
21
22
23

24
25 The IproCo Framework was shown to foster the understanding of collaboration process
26
27 and outcome. So far, it has mostly focused on the role of speaker. Further research has to be
28
29 conducted in order to investigate the processes and influencing factors for the role of addressee.
30
31

32 Practical Implications

33 *Shared applications*

34
35 Our second aim was to provide effective support to the collaborators. In the first
36
37 experiment, we tested technical support measures (the availability of a pictorial or textual shared
38
39 application) to improve the situation and enhance the possibilities for grounding. As the usage
40
41 was not obligatory, only few dyads used the shared application (varying between 6 and 14 dyads
42
43 out of 20). Therefore, no reliable effect could be found for the availability of a shared
44
45 application.
46
47
48
49

50
51 Why did not all dyads use the shared application, if available? Many systems for remote
52
53 collaboration provide shared applications (e.g. Bell, 1997; Erickson & Kellogg, 2003; Roseman &
54
55 Greenberg, 1996; Suthers & Weiner, 1995). However, communication via shared applications
56
57
58
59
60

1
2
3 holds additional costs and has not always been shown to improve the communication process
4
5 (e.g. Ertl, Fischer & Mandl, 2006; Dillenbourg & Traum, 1999; Whittaker et al., 1993). To learn
6
7 about the role of shared applications for supporting individual problem-solving and
8
9 communicative demands, we inspected the rate and the benefit of use on a descriptive level.
10
11 Overall, the costs of usage seemed lower for pictorial shared applications, as more dyads used
12
13 the pictorial compared to the textual shared application. In dependence of the type of task, the
14
15 rate of use of pictorial and shared application differed: More dyads used the pictorial shared
16
17 application in abstract tasks, whereas more dyads used the textual shared application in concrete
18
19 tasks. The possibility for indicative gestures to support the establishment of referential identity –
20
21 a function inherent to the pictorial shared application only – seemed to have been perceived as
22
23 helpful particularly in abstract tasks. On the other hand, the costs of typing descriptions in the
24
25 textual shared application had been lower for concrete tasks. Especially non-interactive dyads
26
27 were willing to invest this additional effort, as they used the textual shared application more
28
29 often.
30
31
32
33
34
35

36 A comparison of dyads that did vs. did not use the shared application revealed a benefit
37
38 of use depending on the condition: Dyads in interactive conditions benefited from the use of the
39
40 pictorial shared application, as they committed fewer feature not mentioned errors and fewer
41
42 name and position of feature errors if they made use of the application. Dyads in non-interactive
43
44 conditions benefited from the use of the textual shared application, as they committed fewer
45
46 feature not mentioned errors and could therefore place more pictures on the correct position.
47
48
49

50 We presented examples from the verbal communication data and the data of the shared
51
52 applications (drawings and text) in order to illustrate the benefit of the shared applications for
53
54 establishing mutual understanding and to show the main type of use: Both shared applications
55
56
57
58
59
60

1
2
3 were mostly used to describe the feature value of each single picture (communicative demand)
4
5 instead of using them to support the individual problem-solving demands, i.e. by supporting
6
7 feature identification or using them as a kind of external group memory (Dillenbourg & Traum,
8
9 1999).
10
11

12
13 Thus, for the design of systems for remote collaboration, our results imply tailoring the
14
15 shared applications to support the communicative demands of a task. This does not necessarily
16
17 mean providing highly task specific shared applications (e.g. Ertl et al., 2006) but the situation
18
19 and task characteristics that influence the communicative demands should be closely taken into
20
21 account. In the following, we exemplarily show how such an analysis using the IproCo-
22
23 Framework can help predicting the need for the availability of shared applications.
24
25

26 27 *When to provide which kind of shared application?* 28

29
30 As stated above, the IproCo Framework can help to clarify task and situation
31
32 characteristics, as well as demands on the collaborators and help to decide if or what kind of
33
34 shared applications should be provided. To exemplify such an analysis, we used IproCo to
35
36 reinterpret the findings from Whittaker and colleagues (1993), who presented results from three
37
38 studies with different tasks solved by remote dyads. In these tasks, the dyads either
39
40 communicated only with an audio link or with an additional shared workspace. In terms of the
41
42 IproCo framework, they varied the *situation characteristic* of shared physical environment with
43
44 a kind of physical copresence through the availability of the shared workspace. To explain the
45
46 differences in terms of the benefit of shared workspaces in the three tasks, the researchers also
47
48 used the concept of demands. However, the terms were not well defined. The three tasks were
49
50 described as follows: an “undemanding” text task, in which the dyad had to generate a list, a
51
52 “demanding” text-editing task, in which the collaborators had to co-annotate a shared text, and a
53
54
55
56
57
58
59
60

1
2
3 graphical design task, with two sets of unshared abstract objects that had to be arranged
4
5 altogether in one plan. The results showed benefits of the shared workspace only in the third
6
7 task, and after some practice for the second task.
8
9

10 In terms of the IproCo framework, the *individual problem-solving demands* were very
11 low in the first task (remembering appropriate words). In the second task, the individual
12 problem-solving demands consisted in text understanding and visual search for specific words or
13 passages. In the third task, by contrast, the collaborators faced multiple individual problem-
14 solving demands, for example those of visual search, object recognition, mental rotation, and
15 memorizing positions of objects.
16
17
18
19
20
21
22
23

24 With regard to the *communicative demands*, the first task was quite undemanding, as the
25 dyads were composed of members from the same work group and talked about familiar topics
26 with perfect common ground. The second task required the establishment of referential identity
27 for the text passage they were currently speaking of. The third task, again, required the most
28 communicational effort: Audience design was crucial because the participants had to describe
29 unfamiliar, abstract objects.
30
31
32
33
34
35
36
37
38

39 As both demands were low in the first task, the additional shared workspace did not show
40 any effect. In the second task, the visual channel enabled pointing and helped to install
41 referential identity, but this added value was not obvious from the start. However, in the third
42 task, the shared workspaces, allowing for physical copresence, had an influence on both kinds of
43 demands: They supported cognitive processes such as mental rotation and memory processes and
44 also communicative demands (see results of our first experiment).
45
46
47
48
49
50
51

52 *Instructional Support Measures*

53
54
55
56
57
58
59
60

Supporting Remote Collaborative Problem-Solving 40

1
2
3 In the second experiment, we did no longer use shared applications to support remote
4 collaboration but aimed to enhance the awareness of the need for audience design and used
5 instructional support measures combining a collaboration model and an analogical collaboration
6 script. Two levels of support were realized, the first one sustaining the problem-solving demands
7 and the second the communicative demands. The analysis of the verbal data indeed showed the
8 intended effect, with fewer problems and errors in conditions with support. Unfortunately, the
9 performance did not reflect the impact of the measures across the board: The complete support
10 conditions did take less time for the description and positioning of the pictures, but the total time
11 on task (including individual and grounding time) was higher and there were no significant
12 differences between the numbers of correctly placed pictures for the three support conditions.
13
14
15
16
17
18
19
20
21
22
23
24
25
26

27 This may be due to the high variance in the number of detected features, a measure which
28 has a crucial influence on the performance. The support was not able to enhance the sensitivity
29 for feature detection, a result which was also found in a change detection task by Williams and
30 Simons (2000).
31
32
33
34
35

36 Rather, we suppose a detrimental effect of the support on the cognitive process of feature
37 detection in interactive conditions: With support, the collaborators searched for features
38 individually, while without support they searched collaboratively. Perhaps searching
39 individually helped to prevent illusions of simplicity (Nickerson, 1999) and also illusions of
40 evidence that can arise if objects are physically co-present (Clark, 1996). However, in non-
41 interactive conditions, the process data suggest a benefit of the support for feature detection.
42
43
44
45
46
47
48
49

50 The results further give rise to the supposition that the process of grounding might also be
51 disturbed if the speaker individually searches for features and decides on names for them.
52
53
54
55
56
57
58
59
60

1
2
3 We implemented only one instructional support measure for both interactive and non-
4
5 interactive conditions to allow for a conservative test of our measures. The support measures
6
7 improved both the process of interactive and of non-interactive communication. Yet, descriptive
8
9 tendencies suggest a different role of the support measures in the two conditions: In interactive
10
11 conditions, dyads without support had no real disadvantage with regard to feature detection
12
13 (percentage of identified features: $\text{mean}_{\text{interactive, without}} = 96.3$; $\text{mean}_{\text{interactive, with}} = 95.7$). In non-
14
15 interactive conditions, speakers had no feedback or help from the addressee with regard to
16
17 feature detection and support was quite helpful: Without support, they missed more features
18
19 (percentage of identified features: $\text{mean}_{\text{non-interactive, without}} = 65.1$; $\text{mean}_{\text{non-interactive, with}} = 77.4$). In
20
21 future research, the effect of more situation-specific support measures should be examined.
22
23
24
25
26

27 All in all, the developed support measures can be used as a starting point to help the
28
29 worker and the mechanic introduced in our example to deal with problem-solving and
30
31 communicative demands and to ensure an effective collaboration when repairing a complex
32
33 engine via remote collaboration. Shared applications, as used in our first experiment, may be
34
35 used to support the communicative demands: They can help to install referential identity, i.e. to
36
37 make sure the workers are using the same words to refer to the same parts of the engine.
38
39 Instructional support measures, as used in our second experiment, may help to improve both the
40
41 problem-solving as well as the communication process. However, our results suggest the need
42
43 for inspecting carefully what types of subtasks are most successfully tackled individually and
44
45 what subtasks have to be solved collaboratively.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning*, 13(3), 175-193. DOI: 10.1046/j.1365-2729.1997.00019.x
2. Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.
3. Barry, C., Morrison, C.M., & Ellis, A.W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology*, 50A, 560-585. DOI: 10.1080/027249897392026.
4. Bell, P. (1997). Using argument representations to make thinking visible for individuals and groups. In R. Hall, N. Miyake, & N. Enyedy (Eds.), *Proceedings of CSCL 1997* (pp. 10-19). Toronto: University of Toronto Press.
5. Bertholet, M., & Spada, H. (2005). Cognitive analysis of a remote collaborative problem-solving task and experimental validation of the resulting demand model. In K. Opwis, & I.-K. Penner (Eds.), *Proceedings of the German Cognitive Science Conference 2005* (pp. 15-20). Basel: Schwabe.
6. Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, 44(2), 123-147.
7. Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22(6), 1482-1493. DOI: 10.1037/0278-7393.22.6.1482.
8. Bromme, R., Hesse, F.W., & Spada, H. (Eds.) (2005). *Barriers and biases in computer-mediated knowledge communication and how they may be overcome*. New York: Springer.

- 1
2
3 9. Bromme, R., Jucks, R., & Wagner, T. (2005). How to refer to 'diabetes'? Language in online
4 health advice. *Applied Cognitive Psychology*, *19*, 569-586. DOI: 10.1002/acp.1099.
5
6
- 7
8 10. Brunel, N., & Ninio, J. (1997). Time to detect the differences between two images presented
9 side by side. *Cognitive Brain Research*, *5*, 273-282. DOI: 10.1016/S0926-
10 6410(97)00003-7.
11
12
- 13
14 11. Clark, H.H. (1996). *Using language*. New York, NY: Cambridge University Press.
15
- 16
17 12. Clark, H.H., & Brennan, S.E. (1991). Grounding in communication. In L.B. Resnick, J.M.
18 Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149).
19 Washington, DC: American Psychological Association. DOI: 10.1037/10096-006.
20
21
22
23
- 24
25 13. Clark, H.H., & Krych, M.A. (2004). Speaking while monitoring addressees for
26 understanding. *Journal of Memory and Language*, *50*, 62-81. DOI:
27 10.1016/j.jml.2003.08.004.
28
29
30
31
- 32
33 14. Clark, H.H., & Murphy, G.L. (1982). Audience design in meaning and reference. In J.-F. Le
34 Ny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287-299). Amsterdam:
35 North-Holland.
36
37
38
- 39
40 15. Clark, H.H., & Schaefer, E.F. (1987). Concealing one's meaning from overhearers. *Journal*
41 *of Memory and Language*, *26*, 209-225. DOI: 10.1016/0749-596X(87)90124-0.
42
43
- 44
45 16. Davis, D., Loftus, E.F., Vanous, S., & Cucciare, M. (2007). 'Unconscious transference' can
46 be an instance of 'change blindness'. *Applied Cognitive Psychology*, published online in
47 Wiley InterScience. DOI: 10.1002/acp.1395.
48
49
- 50
51 17. Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning
52 with instructional design. In P.A. Kirschner (Ed.), *Three worlds of CSCL. Can we*
53 *support CSCL* (pp. 61-91). Heerlen: Open Universiteit Nederland.
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
18. Dillenbourg, P. & Traum, P. (1999). Does a shared screen make a shared understanding? In C. Hoadley, & J. Roschelle (eds), *Proceedings of the Third Computer-Supported Collaborative Learning Conference*, (pp. 127-135). Palo Alto, California.
19. Erickson, T., & Kellogg, W. A. (2003). Knowledge communities: online environments for supporting knowledge management and its social context. In M. Ackermann, V. Pipek, & V. Wulf (eds.), *Sharing expertise: beyond knowledge management* (pp. 299-326). Cambridge, MA: MIT Press.
20. Ertl, B., Fischer, F., & Mandl, H. (2006). Conceptual and socio-cognitive support for collaborative learning in videoconferencing environments. *Computers & Education*, *47*, 298-315.
21. Fussell, S.R., Kraut, R.E., & Siegel, J. (2000). Coordination of communication: effects of shared visual context on collaborative work. *Proceedings of CSCW 2000* (pp. 21-30). NY: ACM Press. DOI: 10.1145/358916.358947.
22. Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E., & Kramer, A.D. (2004). Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction*, *19*, 273-309. DOI: 10.1207/s15327051hci1903_3.
23. Gergle, D., Kraut, R. E., & Fussell, S. R. (2006). The impact of delayed visual feedback on collaborative performance. *Proceedings of CHI 2006* (pp. 1303-1312). NY: ACM Press.
24. Hancock, J.T., & Dunham, P.J. (2001). Language use in computer-mediated communication: The role of coordination devices. *Discourse Processes*, *31*, 91-110. DOI: 10.1207/S15326950dp3101_4.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
25. Hansen, M., Spada, H., & Schneider, A. (2007). Supporting and hindering knowledge communication in a collaborative picture-sorting task. *Journal of Psychology, 215*(4), 226-234. DOI: 10.1027/0044-3409.215.4.228
26. Hermann, F., Rummel, N., & Spada, H. (2001). Solving the case together: The challenge of net-based interdisciplinary collaboration. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *Proceedings of the First European Conference on Computer-Supported Collaborative Learning* (pp. 293-300). Maastricht: McLuhan Institute.
27. Horton, W.S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes, 22*(7), 1-26. DOI: 10.1080/01690960701402933.
28. Horton, W.S., & Gerrig, R.J. (2002). Speakers' experiences and audience design: knowing *when* and knowing *how* to adjust utterances to addressees. *Journal of Memory and Language, 47*, 589-606. DOI: 10.1016/S0749-596X(02)00019-0.
29. Horton, W.S., & Gerrig, R.J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes, 40*, 1-35. DOI: 10.1207/s15326950dp4001_1.
30. Horton, W.S., & Gerrig, R.J. (2005b). The impact of memory demands on audience design during language production. *Cognition, 96*, 127-142. DOI: 10.1016/j.cognition.2004.07.001.
31. Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91-117. DOI: 10.1016/0010-0277(96)81418-1.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
32. Horton, W.S., & Spieler, D.H. (2007). Age-related differences in communication and audience design. *Psychology and Aging*, 22(2), 281-290. DOI: 10.1037/0882-7974.22.2.281.
33. Hupet, M, Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology*, 82(4), 449-462.
34. Isaacs, E.A., & Clark, H.H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116, 26-37. DOI: 10.1037/0096-3445.116.1.26.
35. Krauss, R.M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 14, 343-346. DOI: 10.1037/h0023705.
36. Kraut, R.E., Fussell, S.R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18, 13-49. DOI: 10.1207/S15327051HCI1812_2.
37. Levelt, W.J.M. (2003). Frames of references. In S.C. Levinson (Ed.). *Space in Language and Cognition* (pp. 24-61). Cambridge, UK: Cambridge University Press.
38. McNamara, D.S. (2006). Bringing Cognitive Science into Education, and Back Again: The Value of Interdisciplinary Research. *Cognitive Science*, 30, 605-608. DOI: 10.1207/s15516709cog0000_77.
39. Münzer, S., & Borg, A. (2007). Computer-mediated communication: Synchronicity and compensatory effort. *Applied Cognitive Psychology*, published online in Wiley InterScience. DOI: 10.1002/acp.1387.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
40. Nickerson, R.S. (1999). How we know – and sometimes misjudge – what others know: Imputing one’s one knowledge to others. *Psychological Bulletin*, 125(6), 737-759. DOI: 10.1037/0033-2909.125.6.737.
41. Nückles, M., & Ertelt, A. (2006). The problem of describing a problem: Supporting laypersons in presenting their queries to the internet-based helpdesk. *International Journal of Human-Computer Studies*, 64, 648-669. DOI: 10.1016/j.ijhcs.2006.01.001.
42. O’Donnell, A.M., & Dansereau, D.F. (1992). Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 120-141). New York: Cambridge University Press.
43. Oldfield, R.C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17, 273-281.
44. Olson, G.M., & Olson, J.S. (2000). Distance Matters. *Human-Computer Interaction*, 15, 139-178. DOI: 10.1207/S15327051HCI1523_4.
45. O’Regan, J.K., Deubel, H., Clark, J.J., & Rensink, R.A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, 7, 191-211. DOI: 10.1080/135062800394766.
46. Pearson, P.M., & Schaefer, E.G. (2005). Toupee or not toupee? The role of instructional set, centrality, and relevance in change blindness. *Visual Cognition*, 12(8), 1528-1543. DOI: 10.1080/13506280444000814.
47. Pickering, M.J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169-226. DOI: 10.1017/S0140525X04000056

Supporting Remote Collaborative Problem-Solving 48

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
48. Rensink, R.A. (2002). Change detection. *Annual Review of Psychology*, 53, 245-277. DOI: 10.1146/annurev.psych.53.100901.135125.
49. Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5), 368-373. DOI: 10.1111/j.1467-9280.1997.tb00427.x.
50. Roseman, M., & Greenberg, S. (1996). TeamRooms: Network Places for Collaboration. In Proceedings of ACM CSCW'96 (pp. 325-333). ACM Press.53. Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem-solving in a desktop-videoconferencing setting. *The Journal of the Learning Sciences*, 14(2), 201-241. DOI: 10.1207/s15327809jls1402_2.
51. Schober, M. (1993). Spatial perspective-taking in conversation. *Cognition*, 47, 1-24. DOI: 10.1016/0010-0277(93)90060-9.
52. Scott-Brown, K.C., Baker, M.R., & Orbach, H.S. (2000). Comparison blindness. *Visual Cognition*, 7, 253-267. DOI: 10.1080/135062800394793.
53. Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, 86(2), 420-428. DOI: 10.1037/0033-2909.86.2.420.
54. Simons, D.J., & Levin, D.T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review*, 5, 644-649.
55. Smilek, D., Eastwood, J. D., & Merikle, P. M. (2000). Does unattended information facilitate change detection? *Journal of Experimental Psychology: Human Perception & Performance*, 26(2), 480-487. DOI: 10.1037/0096-1523.26.2.480.
56. Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of*

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Experimental Psychology: Human Learning and Memory*, 6(2), 174-215. DOI:
10.1037/0278-7393.6.2.174.
57. Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making:
biased information sampling during discussion. *Journal of Personality and Social
Psychology*, 48(6), 1467-1478. DOI: 10.1037/0022-3514.48.6.1467.
58. Suthers, D., & Weiner, A. (1995). Groupware for developing critical discussion skills.
Proceedings of CSCL 1995 (pp.341-348). Mahwah, NJ: Lawrence Erlbaum Associates.
59. Tang, J. (1991). Findings from observational studies of collaborative work. *International
Journal of Man-Machine Studies*, 34, 143-160. DOI: 10.1016/0020-7373(91)90039-A.
60. Treisman, A.M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive
Psychology*, 12(1), 97-136. DOI: 10.1016/0010-0285(80)90005-5.
61. Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in perception of objects.
Cognitive Psychology, 14, 107-141.
62. VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47, 513-539.
DOI: 10.1146/annurev.psych.47.1.513.
63. Whittaker, S., Geelhoed, E., & Robinson, E. (1993). Shared workspaces: how do they work
and when are they useful? *International Journal of Man-Machine Studies*, 39, 813-842.
DOI: 10.1006/imms.1993.1085.
64. Whittaker, S., & O'Connell, B. (1997). The role of vision in face-to-face and mediated
communication. In K. Finn, A. Sellen, & S. Wilbur (Eds.), *Video-mediated
communication* (pp.23-49). Mahwah, NJ: Lawrence Erlbaum Associates.
65. Williams, P., & Simons, D.J. (2000). Detecting changes in novel, complex three-dimensional
objects. *Visual Cognition*, 7, 297-322. DOI: 10.1080/135062800394829.

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.

	Type of pictures	
	Concrete	Abstract
Task characteristic	Familiar objects	Patterns/ geometrical figures
Demands		
Individual problem-solving	Search more error-prone	Search less error-prone
Communicative	Easy to describe	Difficult to describe

Supporting Remote Collaborative Problem-Solving 52

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.

Dependent variable	Type of pictures		Mode of communication	
	Concrete	Abstract	Interactive	Restricted
Process measures				
Identified features (percentage)	86.0 (14.2)	90.6 (13.7)	91.1 (12.7)	85.6 (14.7)
		*		*
“Feature not mentioned” errors (frequency)	9.0 (7.3)	3.4 (3.1)	5.3 (4.4)	7.1 (5.7)
		**		
“Name and position of feature” errors (frequency)	.6 (1.1)	2.4 (2.8)	1.2 (1.5)	1.8 (2.3)
		**		*
“Frame of reference” errors (frequency)	1.2 (3.1)	.2 (1.2)	.6 (1.0)	.8 (2.8)
		**		
Repeated picture descriptions (frequency)	22.6 (6.1)	19.7 (3.6)	20.4 (4.6)	21.8 (4.8)
		**		
Warning signals/ clarification questions asked (frequency)	5.2 (4.5)	7.4 (5.5)	9.1 (4.7)	3.5 (3.1)
		**		**
Coordination activities (frequency)	.6 (.9)	1.1 (1.1)	1.4 (1.3)	.6 (.6)
		**		**

Table 2 (continued)

Dependent variable	Type of pictures		Mode of communication	
	Concrete	Abstract	Interactive	Restricted
Performance measures				
Pictures placed in correct position (percentage)	62.2 (26.7)	86.1 (14.4)	79.4 (17.2)	68.9 (22.2)
		**		*
Time needed to complete the task (in sec.)	600 (200)	972 (295)	738 (251)	835 (237)
		**		**

Supporting Remote Collaborative Problem-Solving 54

Table 3

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

Dependent variable	Type of pictures		Amount of support			Mode of comm.	
	Concrete	Abstract	Level 1+2	Level 1	No supp.	Interactive	Non-interactive
Process measures							
Identified features (percentage)	78.8 (18.4)	90.4 (16.8)	86.3 (13.9)	86.8 (16.6)	80.7 (21.4)	95.5 (5.6)	73.3 (17.1)
		**					**
“Feature not mentioned” errors (frequency)	27.1 (14.3)	9.2 (12.6)	14.4 (11.5)	15.9 (10.3)	24.5 (14.6)	12.6 (10.5)	23.7 (13.7)
		**		*			**
Features revealed before description (percentage)	31.3 (32.1)	45.3 (46.0)	67.8 (30.6)	31.8 (37.9)	15.3 (25.8)	50.7 (39.3)	25.9 (35.7)
		**		**			**
“Name and position of feature” errors (frequency)	0 (0)	1.2 (1.8)	.5 (.8)	.7 (1.2)	.7 (.8)	.4 (.7)	.8 (1.0)
		**					

Table 3 (continued)

Dependent variable	Concrete	Abstract	Level 1+2	Level 1	No supp.	Interactive	Non- interactive
“Frame of reference” errors (frequency)	2.7 (4.8)	0 (0)	.5 (.8)	1.4 (3.2)	2.1 (2.4)	1.4 (2.3)	1.3 (2.2)
	**						
Irrelevant features (frequency)	13.7 (13.0)	3.1 (7.0)	6.9 (9.4)	8.3 (9.4)	9.9 (6.1)	8.0 (9.9)	8.7 (10.2)
	*						
Complicated descriptions (frequency)	.4 (1.4)	2.0 (3.2)	.6 (1.3)	1.2 (1.6)	1.8 (2.5)	.8 (1.8)	1.6 (2.8)
	**						
Time period measures							
Individual time (in sec.)	358 (334)	550 (429)	705 (398)	411 (201)	155 (149)	347 (278)	561 (432)
	**			**		**	
Information time (in sec.)	73 (89)	103 (115)	156 (102)	55 (72)	52 (96)	118 (104)	58 (92)
	**			**		*	

Supporting Remote Collaborative Problem-Solving 56

Table 3 (continued)

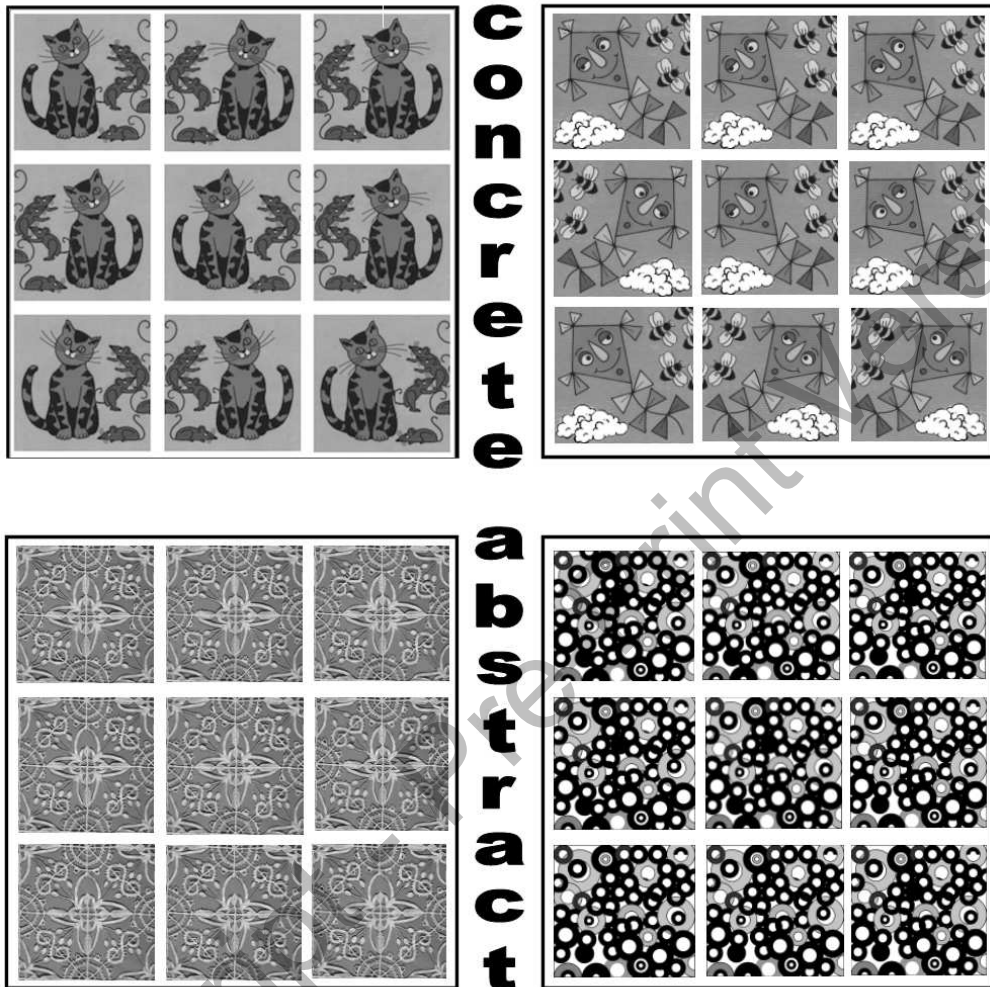
Dependent variable	Concrete	Abstract	Level 1+2	Level 1	No supp.	Interactive	Non- interactive
Performance measures							
Pictures placed in correct position (percentage)	32.8 (41.7)	72.2 (40.0)	51.7 (40.0)	57.2 (36.1)	49.4 (47.2)	80 (21.7)	25.6 (28.9)
		**					**
Time needed for description and positioning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

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).

Figure 1 TOP



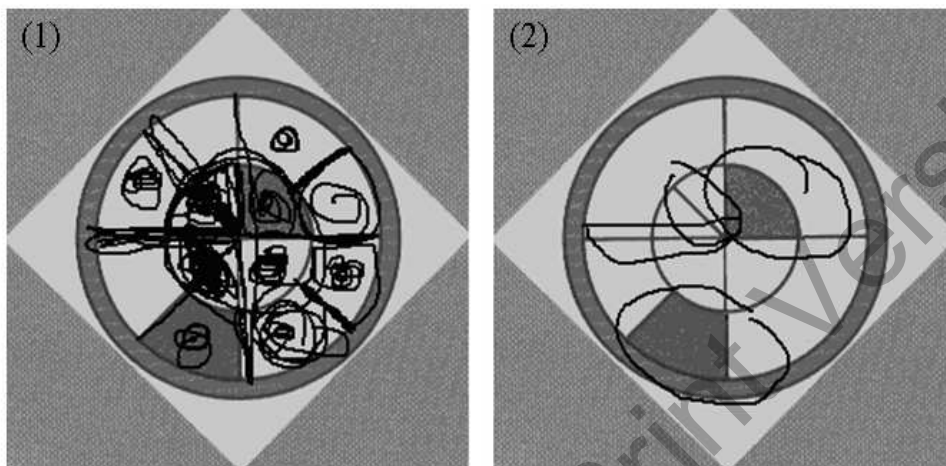
349x370mm (72 x 72 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Manuscript Central

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 2 TOP



246x141mm (72 x 72 DPI)

Manuscript - Pre Review