Playing Pictionary: An Exploration in Human-Computer Interaction

Eric Brochu, Peter Carbonetto, Karyn Moffatt and Pinar Muyan
University of British Columbia - CPSC 544 Term Project

Abstract

Due to the inability to successfully reproduce face-to-face conditions, computer-mediated multiparty activities suffer from inferior communication. It was therefore desired that a study be performed in order to determine whether methods exist to mitigate the negative effects of distance on collaborative communication tasks. A study was performed examining the effects of distance on subjects' performance on the task of playing Pictionary. Two numerical results were recorded from the sessions – the time taken to get the correct answer, and the number of vocalizations before getting the correct answer. Although, no significant results were found in the analysis of the means, trends indicate that subjects were faster in the face-to-face setting. It is suspected that the insignificance of the results is due in large part to the vast variance between words for the time taken to guess correctly. In the analysis of the frequency of vocalizations, there was a significant result of experimental setup. Subjects made more vocalizations in the face-to-face setting. It also was observed that group interaction and dynamic was pivotal to the team performance.

Background

This project was originally conceived as a means to validate the distance art therapy software developed by Kate Collie et al. Through the course of our discussions, we conceded the need to narrow the focus of our study so that we could effectively investigate a dependent variable under strict experimental conditions. We thus designed and conducted an experiment using an online version of the game Pictionary to isolate and measure the variable.

Purpose of the Experiment

Face-to-face signifiers are an integral part of both formal and informal communication. Non-verbal cues include the manner in which words are spoken (paralinguistic speech), facial expressions, eye contact, cueing using gaze, posture, and physical gestures.

Many non-verbal cues are redundantly encoded in verbal communication, and many such cues may be dropped with little or no impact on the overall ability of a group to communicate. For example, it has been suggested that a speaker yielding the floor could use any of eight distinct verbal and non-verbal cues to convey that intent [9]. Nevertheless, there exist many non-verbal cues that are used to convey information distinct from and/or complimentary to verbal communication. Pointing is used to guide attention to items on a diagram or towards members of a group. Eye gaze is another such example – it is a subtle mechanism for directing participants' attention, indicating interest to a speaker and achieving deixis.

Due to the inability to successfully reproduce face-to-face conditions, computer-mediated multiparty activities often suffer by comparison. Face-to-face interaction is not only important for conveying non-verbal cues, but also for building confidence and establishing trust [8]. The members of a group impulsively gravitate towards establishing trust with each other in order to complete a group task in a face-to-face context, but this is not necessarily true in a computer-mediated context. Video conferencing systems try to overcome this communication deficit but most only make a small step to matching the conditions of face-to-face interaction. In practice, video-conferencing systems hold little, if any, improvement over their audio-only counterparts because video does not provide spatial relative information, and therefore does not permit participants to visually direct their communication to group members. In [8] Nohria and Eccles concisely claim, "effectiveness of electronic network will depend on an underlying network of social relationships, based on face-to-face relationships."

Current videoconferencing systems are also unable to properly convey a sense of physical awareness, including conditions of a physical workspace. Information available in a physical workspace allows participants to maintain a context-rich representation of others group members' locations, activities, intentions and other intangibles that enable individuals to mentally model their fellow group members. A sense of physical awareness of others allows people to work more effectively [5, 7]. This awareness of physicality -- important for social interaction in communication technologies -- is usually referred to as *tele-presence*.

A great deal of research has been conducted on improving computer-mediated multiparty communication by supporting interfaces for tele-presence and better simulating face-to-face conditions [4, 6, 9]. In this experiment, we call into question the very motivation for all this research. Is there a way to compensate for the lack of physical presence in computer-mediated group tasks without directly recreating natural group communication conditions? For a simple group task, are the visual cues from face-to-face relations replaceable without affecting task performance?

The intention of our experiment was to investigate the importance of face-to-face interaction and physical presence on communication in a simple group task involving active drawing, and to ascertain whether members of a computer-mediated group can compensate for the lack of non-verbal cues by mutually setting guidelines for interaction. "Simple" may be a bit of a misnomer because all group tasks are inherently complex. What we mean to imply here by "simple" is that the group task does not implicate relatively advanced social and psychological multiparty interaction, and that the task is almost entirely contained within the scope of the experimental conditions.

Hypothesis

Previous research has suggested that the social benefits of face-to-face contact are irreproducible by artificial means. Face-to-face interaction allows us to make eye contact, observe other's myriad facial expressions and gestures.

An important element of any collaborative task is trust. Trust must be maintained between all the members of the group in order to properly cooperate, resolve dilemmas and ultimately complete the task successfully. Evidence shows that face-to-face contact is a major ingredient in building and maintaining trust. In an electronic context without face-to-face contact, trust may break down more easily, especially when challenges arise, either internal or external. More encouragingly, a study by Elena Rocco shows that trust can be achieved when team members have initial face-to-face contact. Early establishment of guidelines shapes cooperative attitudes for later improvement in team dynamics [8].

Mutual adjustment is also a very delicate resource rooted in face-to-face relations. Without face-to-face contact, a group might have trouble adjusting communication to situations that warrant it [8]. If a group needs to adjust to compensate for lack of visual cues in communication, there is a question as to whether the computer-mediated context would impose additional obstacles in accomplishing that objective.

Non-verbal signifiers, such as gaze cueing and pointing, are needed for shifting the focus of attention from interpersonal spaces to the shared workspace and vice-versa. When talking about a drawing task, the shared workspace is the drawing surface. Non-verbal cues aid in directing attention towards or away from either the drawing or the discussion between the members of the group.

We are not implying that visual cues are irreplaceable. In a context lacking visual communication, verbal cues can often substitute for non-verbal ones. For example, a speaker can focus the attention of the group to an individual by saying his or her name in a certain manner or, more explicitly, by saying, "let's listen to John now."

Our hypothesis is as follows: *It is given that there is a definite impediment to performing goal-oriented group tasks when separated from, but in communication with other members of the group. However, given a priori opportunity to create mental models of the other members of a group and sufficiently incomplex cooperative task to permit recognition of communication problems within the scope of the experiment, we predict that the negative effects of separation can be substantially mitigated.*

Experimental Setup

After considering various alternatives, we settled on the use of the parlor game Pictionary as a means by which to test our hypothesis (see *Appendix A:* Rules to Pictionary and *Appendix A:* Rules to Pictionary for a description of the game). We divided our subjects into teams of three to four and had them play a modified version of Pictionary using online whiteboard technology. Each group played two matches – one in which they were all in the same room, and one in which they were all in separate rooms.

We conducted the experiment in two sessions. Each session involved two groups and two setups. While the first group (hereafter identified as Group A) was playing the game in the face-to-face setup (F2F), the second (Group B) was playing in the distance setup. After both groups finished, they switched setups.

Match	Group A	Group B
First	Face-to-face (F2F)	Distance
Second	Distance	Face-to-face (F2F)

Table 1: Trial Cardinality

Before the experiment, we pre-selected two ordered lists of 10 words each. We labeled the lists *face-to-face* and *distance*. First, we selected 20 words from the Pictionary board game then we ordered them randomly into the two lists by flipping a coin. The resultant word lists were:

	Face-to-face	Distance
1	Cut corner	Forget
2	Party animal	Sip
3	Earthquake	Swiss cheese
4	Toxic waste	Hop
5	Caviar	Tickle
6	Toy	Amazon River
7	Blush	Girlfriend
8	Off	Godzilla

9	Barbie	Corn flakes
10	Ink	Pluck

Table 2: Word Lists

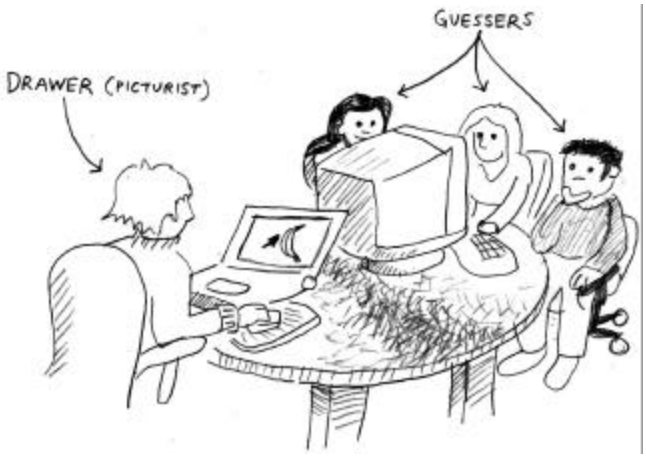

Figure 1: Group playing Pictionary in face-to-face scenario.

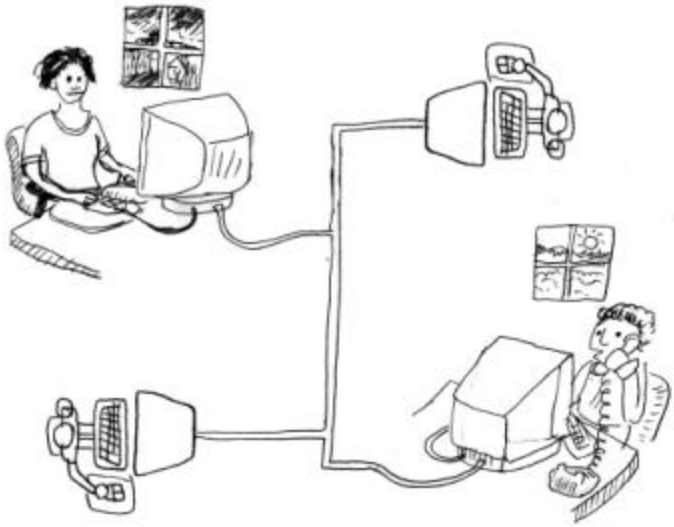

Figure 2: Group playing Pictionary in distance scenario

.

The drawing medium was the computer. In the F2F trials, the guessers all shared a single computer between them. The computers were positioned so that the guessers and the picturist could physically interact with each other without much hindrance. See *Figure 1* for an example of the face-to-face experimental setup.

In the distance situation, all the participants were in separate rooms with their own computers. See *Figure 2* for an illustration of the distance experiment setup.

Equipment

The following sections describe the equipment used in performing this study.

Software

We used both Windows 2000 and RedHat Linux workstations. All the workstations were connected to a high-speed Ethernet network, which permitted us to run the drawing software with minimal lag.

We used a web conferencing Java applet for playing Pictionary in both the face-to-face and distance scenarios. This allowed us to control for variables such as network lag, computer interface and technical challenges associated with the particulars of the software. Since the software was written in Java, and running as a Netscape applet, we were able to achieve platform

transparency. We scheduled and ran the web conference using PlaceWare Web Conferencing demo software, available online at http://www.placeware.com.

Figure 3: PlaceWare Videoconferencing software.

PlaceWare has two modes, one for the presenter and one for the viewers. In our case, the designated drawer was logged into the web conference as a presenter, and the guessers were logged in as viewers. The presenter mode has more functionality. The presenter can create and draw new slides. The guessers can see the slides but cannot alter them.

Before starting the game, we introduced the participants to the features of the software and gave them a chance to familiarize themselves with the interface. *Figure 3* shows the PlaceWare interface. There is a freehand drawing tool, a straight line drawing tool, a square-drawing tool, and "stamps" for placing arrows, checkmarks and crosses. There is also an eraser, a clear button that deletes the entire drawing space, a textbox tool and a special tool for showing cursor position to all participants as a red dot on their screens. Other than the textbox tool, the drawers could use all the tools freely.

The freehand tool was by far most often used. In the screenshot above, the freehand tool on the top-right portion of the screen is currently highlighted. Some – but not all – teams established conventions for the stamps and the cursor position tool. No team used the square tool. During the experiment we found that the subjects used the tools in interesting and novel ways.

The cursor position tool was a useful feature for playing Pictionary – it allowed the picturist to direct the guessers' attention to certain regions of the drawing. And indeed, during the experiment the participants quickly took to the pointer tool since it was fairly easy to use and effective. The clear button allowed the picturist to refresh the drawing space. This was an intuitive feature since it was the equivalent of throwing away the piece of paper when playing the game normally.

At the end of each word, we asked the picturist to click on the "New Slide" button to save the final drawing.

Overall, the software setup worked well on our workstations, some of which were several years old. We did experience a technical failures -- during the course of the experiment, we discovered that the PlaceWare Java applet crashed whenever the users tried to resize the drawing window. After discovering this flaw we warned subsequent users not to resize the window.

Video camera setup

Ideally, we would have liked to record the experiment using a total of six video cameras, with one on each of the picturists, one on each of the guessers in the distance setup, and one on the group of guessers in the F2F setup. Unfortunately, this was not practical. Instead, we used two cameras: a Hi-8 camera and a digital video camera, one for each of the setups, which were running simultaneously. In the first session, the Hi-8 video camera recorded the F2F situation and the digital video camera recorded the drawer in the distance case. The digital recording was easy to edit on a Macintosh using iMovie software, so it was ideal for the purposes of our presentation. For the second session, we switched the cameras. This enabled us to combine footage from both the face-to-face and distance situations in our presentation.

In the distance scenario, we used an additional laptop to log in to the web conference. We positioned the laptop so that we could record the drawings in addition to the actions and facial expressions of the picturist. In the face-to-face sessions, we were able to train the camera on the drawer in one session, and the guessers in the other session – never both at the same time, unfortunately, due to the size of the room.

Audio setup

We attached microphones to the video cameras and placed them on the table. We also recorded audio using tape recorders. In the face-to-face situation, we placed the tape recorder on the table near the guessers. This setup provided additional audio recording for ease of subsequent analysis. In the distance scenario, we attached the tape recorder to the telephone, and this resulted in a superior record of the participants' discourse.

Timing of subjects

We used stopwatches to measure the amount of time it took a team to guess the correct word. We had a facilitator time the subjects for each group conducted in parallel. Our error of measurement was approximately ±1 second.

Measurement

Our measure of performance was the amount of time it took a group to correctly guess the word being drawn. We also recorded the number of incorrect guesses for each word. At the end of each session, we had the subjects fill out a survey. The survey and responses are detailed in Appendix C.

Controlling the Variance

As detailed in our hypothesis, the primary goal of this experiment was to measure performance of a collaborative task in two different situations, in which the independent variable was the ability to communicate through face-to-face interaction, in order to measure the effect of the absence of non-verbal cues on communication through active drawing.

We took several measures to mitigate the effects of factors that might introduce inconsistencies between the proximal and separated groups. We selected subjects whose first language was English or, at worst, those that were highly fluent in English. Since Pictionary is a strongly language-dependent game and many of the associations are culturally rooted, participants whose first language was not English would be at a disadvantage. The skill of the drawer was difficult, if not impossible, to control – there was no way to ensure that the drawers would have similar proficiency in conveying their thoughts visually.

We made sure that each picturist was confident, relatively experienced in playing Pictionary, adept at using the mouse for drawing and, most importantly, comfortable in his or her role as picturist. If the picturist were not comfortable, his or her disposition would negatively affect the overall confidence of the team. In the end, we found the best way to achieve this, given the circumstances, was to let the team elect a drawer from amongst themselves. In every circumstance, there was at least one person in the group who showed enthusiasm in taking up the role as picturist.

Overall, we attempted to make the computer environment as similar as possible in the distance and F2F scenarios. The picturist had to use the computer interface in either situation. We made sure network latency was not a factor by using at least two computers connected over the network in both cases. Even though the guessers were in the same room as the picturist in the F2F situation, the monitors were situated such that they could only see their own. All groups drew the same words in the same situations.

Observations

Statistical Results

The resultant word guessing times of our experiment were as follows:

Word	Group 1A	Group 1B	Group 2A	Group 2B
Cut corner	26	22	84	35
Party animal	38	35	39	42
Earthquake	66	27	-1	-1
Toxic waste	16	40	-1	41
Caviar	11	16	94	77
Toy	48	-1	60	-1
Blush	14	9	20	101
Off	8	11	27	15
Barbie	94	28	-1	93
Ink	10	35	40	26
Mean	*33.10*	*34.30*	*72.40*	*67.00*
Standard deviation	*28.67*	*31.87*	*40.17*	*39.81*

Table 3:Time results of face-to-face sessions

Word	Group 1A	Group 1B	Group 2A	Group 2B
Forget	-1	90	-1	-1
Sip	14	109	46	*
Swiss cheese	5	8	13	31
Hop	45	68	-1	46
Tickle	79	-1	-1	12
Amazon river	49	67	58	-1
Girl friend	-1	18	54	-1
Godzilla	15	30	28	-1
Cornflakes	72	18	39	49
Pluck	34	62	-1	103
Mean	*55.30*	*59.00*	*71.80*	*80.11*
Standard deviation	*41.73*	*39.64*	*43.36*	*44.82*

Table 4: Time results of distance sessions

All times are in seconds.

-1 *Group failed to guess word within 120 seconds (a value of 120 was used in the calculation of the mean).*

* *Invalid results to due technical difficulties.*

Figure 4 below shows each group's mean times for both experimental setups. As the graph shows, each group – with the exception of group A2 – completed the task faster in the face to face setting than in the distance setting.

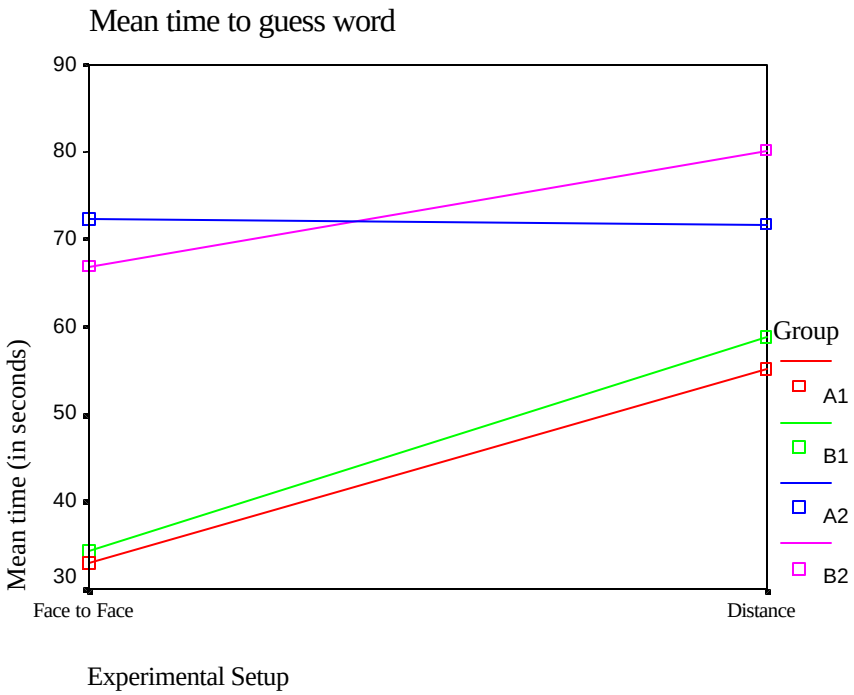

Figure 4: Mean time to guess correct answer versus experimental setup – by group

Figure 5 shows the overall mean time for each experimental setup. Again, groups took less time to guess the correct word in the face-to-face setup than in the distance.

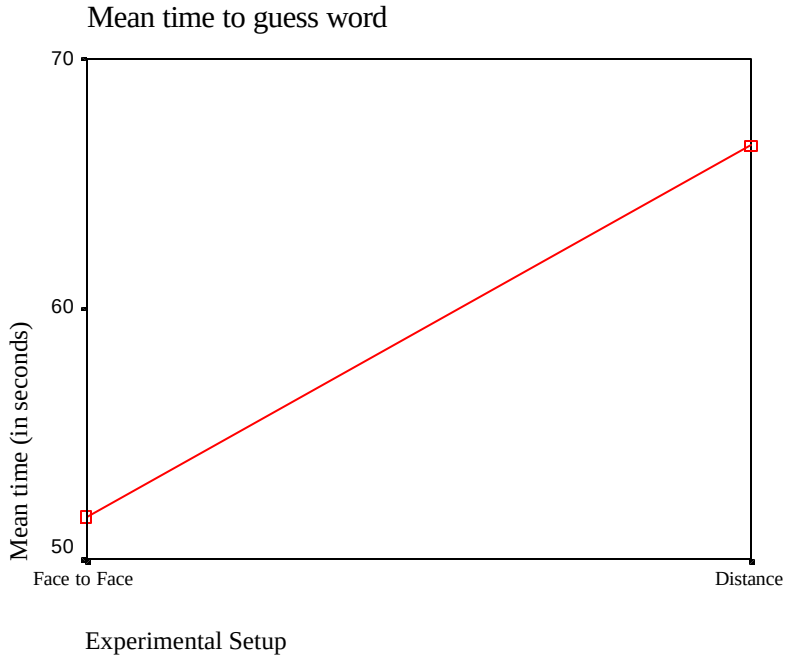

Figure 5: Mean time to guess answer versus experimental setup - overall result

Examining the interaction of order with experimental setup gives the results shown in *Figure 6* below. This graph has a few interesting implications. First, groups did better (compared to the other group in the same setup) in whichever setup they did second. That is, group B did better than group A in the face-to-face setup, and group A did better than group B in the distance setting. This is mostly likely attributable to a practice effect.

More interesting is that the magnitude of the advantage group A had over group B in the distance setting is greater than the magnitude of the advantage group B had over group A in the face-to-face setting. Although not statistically significant, this difference may be due to an effect of having the face-to-face interaction prior to the distance task over and above mere practice effect. This effect would be consistent with research that has found that face-to-face interaction allows subjects to develop a mental model of the members of the group that then helps them to perform in group tasks in a subsequent distance environment.

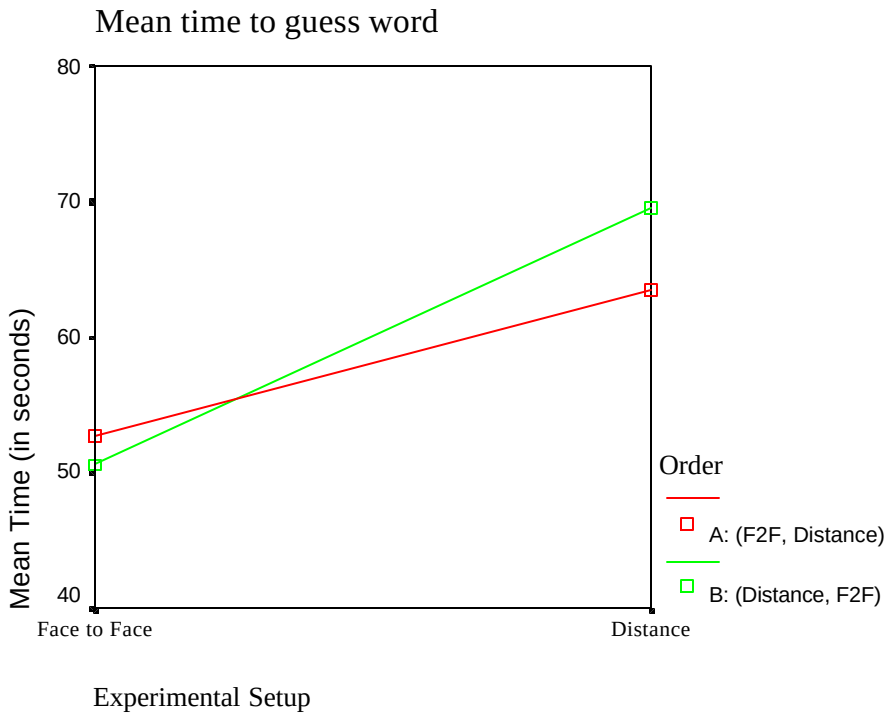

Figure 6: Interaction of Experimental Setup and Order on Mean time to guess word

These results, while they are suggestive of a difference in performance between face-to-face and distance sessions, do not lead us to any statistically significant conclusions (see *Appendix D:* Detailed Statistical Analysis for the details of the statistical analysis). This is due in particular to the very large standard deviation associated with the mean times to guess a word.

We can largely attribute a poor standard deviation to the nature of the task we were observing coupled with the small number of trials. One factor is the difficulty in achieving a uniform distribution of values for any word, or for one group across a series of words. There was both a large within word group and within treatment group variance. Although all teams guessed most

words within 120 seconds, there was no predictability in how long it would take for any single word. Likewise, even for groups in the same treatment format, there was no agreement on how long it would take to guess any particular word. Furthermore, the time values for guessing each word are bounded by our imposed time limit of 120 seconds. While the upper bound could, in theory, be removed to give a "true" value for the amount of time required to guess the word, in our experience, we feel this would not give a more meaningful value, because as time goes on without the guessers correctly identifying the drawing, frustration sets in on the parts of all group members, and the number and quality of guesses decreases.

With a larger subject pool more passes could be performed on each word. In fact, we could run such an analysis in a preliminary study allowing us to determine a difficulty factor for each candidate word allowing us to adjust the individual times of the results to account for difficulty. Alternatively we could use a series of words of similar difficulty to compose the set of guessed words for the experiment.

We also counted the number of incorrect guesses of each word made before the drawing was correctly identified. The results are as follows:

Face to Face	**Session 1**						**Session 2**					
	Group A			**Group B**			**Group A**			**Group B**		
Words	**Times T**	**Guesses NI**	**Freq NI/T**	**Times T**	**Guesses NI**	**Freq NI/T**	**Times T**	**Guesses NI**	**Freq NI/T**	**Times T**	**Guesses NI**	**Freq NI/T**
Cut Corner	26	13	0.50	22	11	0.50	84	22	0.26	35	20	0.57
Party Animal	38	19	0.50	35	18	0.51	39	11	0.28	42	42	1.00
Earthquake	66	29	0.44	27	19	0.70	120	30	0.25	120	11	0.09
Toxic Waste	16	6	0.38	40	24	0.60	120	27	0.23	41	30	0.73
Caviar	11	10	0.91	16	10	0.63	94	18	0.19	77	40	0.52
Toy	48	18	0.38	120	40	0.33	60	21	0.35	120	30	0.25
Blush	14	5	0.36	9	4	0.44	20	4	0.20	101	10	0.10
Off	8	2	0.25	11	21	1.91	27	5	0.19	15	7	0.47
Barbie	94	30	0.32	28	12	0.43	120	18	0.15	93	31	0.33
Ink	10	4	0.40	35	18	0.51	40	7	0.18	26	7	0.27
Mean	33.10		0.44	34.30		0.66	72.4		0.23	67.00		0.43
Overall – Face to Face												
Mean T	51.70											
Mean NI/T	0.44											

Table 5: Mean frequency of vocalizations for each group – Face-to-face

Distance	Session 1						Session 2					
	Group A			Group B			Group A			Group B		
Words	Times T	Guesses NI	Freq NI/T	Times T	Guesses NI	Freq NI/T	Times T	Guesses NI	Freq NI/T	Times T	Guesses NI	Freq NI/T
Forget	120	***		90	22	0.24	120	24	0.20	120	20	0.17
Sip	14	***		109	22	0.20	46	14	0.30		*	
Swiss Cheese	5	0	0.00	8	3	0.38	13	5	0.38	31	8	0.26
Hop	45	2	0.04	68	6	0.09	120	24	0.20	46	16	0.35
Tickle	79	***		120	32	0.27	120	22	0.18	12	6	0.50
Amazon River	49	***		67	17	0.25	58	14	0.24	120	40	0.33
Girlfriend	120	***		18	***		54	15	0.28	120	32	0.27
Godzilla	15	***		30	***		28	8	0.29	120	28	0.23
Cornflakes	72	26	0.36	18	***		39	11	0.28	49	16	0.33
Pluck	34	17	0.50	62	9	0.15	120	22	0.18	103	25	0.24
Mean	55.30		0.23	59.00		0.23	71.80		0.25	80.11		0.30
Overall - Distance												
Mean T	66.55											
Mean NI/T	0.25											

Table 6: Mean frequency of vocalizations for each group – Distance

* *Invalid results to due technical difficulties.*
*** *Unknown count due to technical difficulties.*
T *Time it took to get the correct answer*
NI *Number of incorrect answers.*
NI/T *Frequency of the vocalizations*

An Anova was run on the above numbers to determine the effect experimental setup on the frequency of vocalizations. *Table 7* shows the numerical results of this analysis, and *Figure 7* illustrates these results in a graph.

Tests of Between-Subjects Effects

Dependent Variable: Frequency of Vocalizations

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SETUP	.580	1	.580	9.403	.003
Error	4.193	68	6.166E-02		
Total	4.773	69			

Table 7: Results of the Anova analysis of setup on frequency of vocalizations

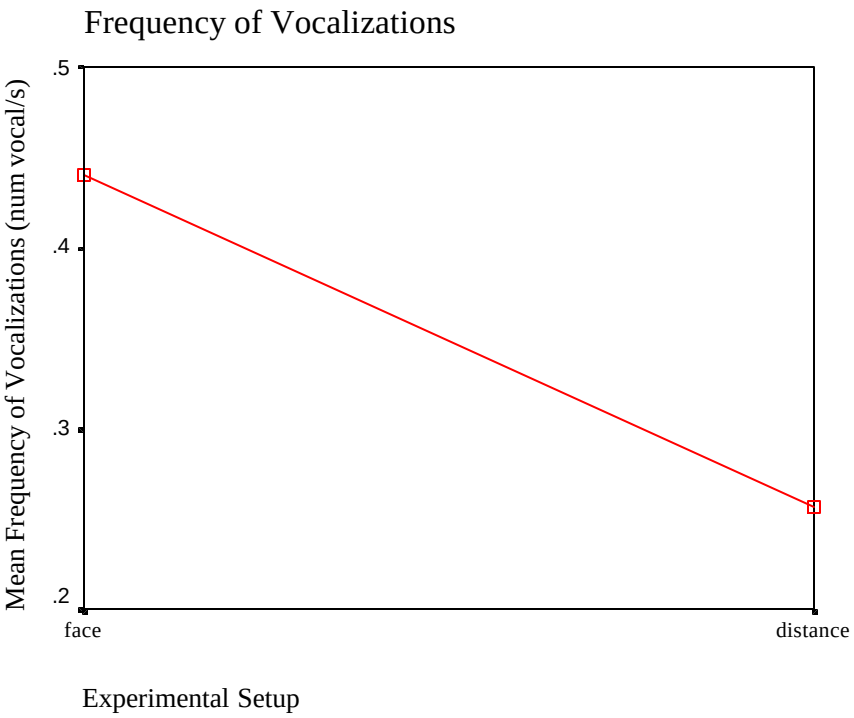

Figure 7: Frequency of Vocalization versus Experimental Setup

The results are as follows:

There is a significant effect of setup, F (1,29) = 8.07, p <= 0.01, on the frequency of guesses vocalized. In the face-to-face setting participants offered guesses more often than they offered when in the distance setting.

We can postulate, albeit with little degree of certainty, that the reason for this effect was the increased interaction between guessers in the face-to-face situation.

Social Observations

Most participants reported that they had a good time and enjoyed the process. This is noteworthy as it had been suggested that the whole experience might have been frustrating for the participants during the distance part of the experiment. In actual fact, the level of frustration reported was very low.

The groups reported feeling at ease with one another and acted as a unit, making explicit references to being part of a "team". This element of group cohesion was most likely aided by the initial social interaction at the beginning of the experiment and was reinforced between words through the course of the session. In the Thursday session, we experienced several minutes of technical difficulty with the software while a group was playing Pictionary in the distance scenario. Rather than becoming frustrated and impatient, the group participants used the delay in the experiment to their advantage by interacting and getting to know each other better.

They started asking each other where they grew up and so on. Also, the small break when the groups switched from F2F to distance and vice-versa appeared to be especially beneficial for the participants from the B-groups; when they reconvened in the same room, they animatedly discussed their performance and the strategies they used. Several participants commented on the fact that the intermittent social interaction when the group task was not being performed was very important for group cohesion.

Relating to the above observation, we note that the A-groups performed better than the B-groups. While they did not perform better to a statistically significant degree, this could be very interesting source for further study. The A-groups performed the task in the face-to-face scenario first. This allowed them to construct a strong mental model of the group members. The B-groups did not have such an opportunity – they did interact socially, but did not have a chance to work cooperatively with each other on the collaborative task in a face-to-face setting. This is what we had hoped to show in our hypothesis when we stated that one of the preconditions for sufficient performance in the group task was the opportunity to create a mental model of the group.

From one group to another, there was a great deal of variance in the level of cooperation and nature of team interaction. We believe that this was one bigger factors relating to a high level of variance in our data, and thus lack of significant results. For example, in one of our four groups, the drawer was a natural leader and was clearly comfortable in directing and promoting cooperation among the members of the team. Other groups appeared to have more difficulty achieving unity or consensus in initiatives because they did not have that important element of leadership. This was strongly evident while observing the groups interact; the drawer-as-leader clearly had a natural role in forming a cooperative attitude for the team.

There are many more issues related to the effectiveness of group collaboration, and we can't possibly detail them all within the scope of this project. We can say that, in the future, this study would be greatly improved by forming more compatible groups. By doing so, we might not control for differences in group dynamics but at least we would ensure a closer level of comfort and ease in communication in each group. In this study, we have to concede that we underestimated the impact of group dynamics.

Frustration and Tools

Before doing the study, we were worried that the users might become frustrated or have difficulty adapting to the drawing interface, especially in a situation where the drawers are expected to sketch things as quickly as possible. In fact, we found that the subjects climbed the learning curve very rapidly and even showed a great deal of creativity when using the interface. Most of the drawers learned how to use the arrow tool to quickly point to portions of their drawings, without instruction from the facilitators. They also demonstrated a great deal of ingenuity with the interface, especially with regards to establishing guidelines for communication. This is discussed in more detail below. Not surprisingly, some drawers still reported that the mouse was difficult to use.

Methods of Compensation

There were some interesting empirical results regarding compensation for the lack of visual cues. Several groups created their own conventions for the game. None of these compensatory measures resulted in significantly improved performance, but these observations are interesting in and of themselves.

Of the four groups, three established conventions for communication either before or during the distance session. For example, one group used the checkmark to indicate that the guessers were on the right track and the cross to indicate that they were going in the wrong direction. Another group established near the end of their session that the guessers often needed to make a slight modification to a suggested word to get the correct answer (e.g. from "drifting" to "drifter"). The team agreed upon a cross mark with a circle around it as an appropriate tool to convey the message.

There was no indication that any of these measures could compensate for the lack of visual cues, but it does show that most groups are willing and motivated to cooperate in a distance setting to form such communication guidelines. It is also interesting to note the one team that did not achieve an agreement for mutually-defined visual symbols also unanimously believed they could perform just as well as in the face-to-face scenario – though their performance suggested otherwise.

In the distance scenario, we often observed drawers reflexively pointing to the screen or making gestures, even though none of these actions were effectual.

Future Improvements

In the course of our experiment, we became aware of the fact that many of our conscious design choices may have negatively impacted the conclusiveness of our study. For future investigations of this hypothesis, some changes to the experiment design should be considered. Many suggestions have already been discussed under "observations", but there are several others that our experiences brought to our attention.

For the F2F setup, our original design suggested that the drawer and guessers were to be placed directly opposite each other. In practice, the physical limitations of the room forced us to deviate from this design. The monitor and the guessers were at a 90-degree angle to one another, impeding the drawer's ability to switch focus between the interpersonal space and the shared workspace.

Improved analysis of group interaction would have been possible if the camera was set up to capture to capture the entire team, instead of just the drawer or just the guessers. Again, a significantly larger room in which to conduct the experiment would have enabled us to do this.

As mentioned previously, it may have been better to recruit groups formed outside the experimental context. In other words, we could consider taking on volunteer teams rather than

individuals. This may have resulted in more natural groups, in which a bond of trust had already been formed outside the experiment. Alternatively, “icebreakers” could perhaps have been conducted early on in the session to increase the comfort level of the participants with their teammates. Although this would add a significant amount of time to the experiment, it may help create more cohesive teams.

Conclusion

It must be first stated that we failed to make any statistically significant conclusions about our hypothesis. We cannot conclude that the negative effects of separation can be substantially mitigated. However, the experiment did lead to some interesting observations and provide inspiration for future inquiry.

We believe our hypothesis could be more effectively verified through the improvements we proposed above. In brief, we could improve the study by increasing the number of subjects and controlling for group dynamics, social interaction and variability among the words.

Additionally, we found promising evidence showing that our hypothesis is in the right direction. Observational evidence supports the argument that initial face-to-face social interaction is an important element in establishing trust and mental models of the group members for future collaboration. This was demonstrated by the observation that the A-groups performed better than B-groups on the distance task, and to a greater extent that practice effect would predict. Also, given a priori mental models, we have anecdotal evidence that participants are willing to collaboratively overcome the deficiencies of a computer-mediated interface in the distance scenario in order to improve their performance in an incomplex group task.

References

1. Kevin Baker, Saul Greenberg, Carl Gutwin. "Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration." *Proceedings of the 8th IFIP Working Conference on Engineering for Human-Computer Interaction,* May 2001.
2. Kate Collie and David Cubranic. "An art therapy solution to a telehealth problem." *Art Therapy: The Journal of the American Art Therapy Association*, Vol. 16, No. 4, pp. 186-193
3. Davor Cubranic, Kellog S. Booth and Kate Collie. "Computer Support for Distance Art Therapy." *Summary of ACM SIGCHI Conference on Human Factors in Computing,* 1998, pp. 277-278.
4. Paul Dourish and Sara Bly. "Portholes: Supporting Awareness in a Distributed Work Group." *Proceedings of CHI Human Factors in Computing Systems,* May 1992, pp. 541-547.
5. Carl Gutwin, Saul Greenberg and Mark Roseman. "Workspace Awareness in Real-Time Distributed Groupware: Framework, Widgets, and Evaluation." *CHI Conference Companion,* 1996.
6. Hiroshi Ishii, Minoru Kobayashi and Jonathan Grudin. "Integration of Inter-Personal Space and Shared Workspace: ClearBoard Design and Experiments." *Proceedings of Conference on Computer-Supported Cooperative Work*, November 1992, pp. 33-42.
7. Eric Paulos and John Canny. "Ubiquitous tele-embodiment applications and implications." *International Journal of Human-Computer Studies,* Vol. 46, No. 6, 1997, pp. 861-877.
8. Elena Rocco. "Trust Breaks Down in Electronic Contexts but Can Be Repaired by Some Initial Face-to-Face Contact." *Proceedings of Conference on Human Factors and Computing Systems,* 1998, pp. 496-502.
9. Roel Vertegaal. "The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration." *CHI Conference on Human Factors in Computing Systems,* 1999, pp. 294-301.

Appendix A: Rules to Pictionary

EQUIPMENT
Game board, one-minute timer, 500 word cards, die, 4 markers, 4 category cards, 4 pads of paper, and 4 pencils.

THE OBJECT
To identify through sketched clues as many words as necessary to advance to the finish square, and correctly identify the final word. Sketches may NOT include letters, numbers or the # symbol.

PREPARATION
Place the timer and card box so all players have access to them. Divide equally into teams of two to four (see instructions on number of players). Provide each team with pad, pencil, category card and marker. Place marker in the start square on the board. Each team selects a picturist, one who will sketch clues for the first word. Roll the die to determine the order of play. The highest roller starts.

THE PLAY
All markers rest in the P square to start, so the word in the P category is in play. The die is NOT rolled to advance at the start.

P Person/Place/Animal (or related characteristics)
O Object (things that can be touched or seen)
A Action (things that can be performed; events)
D..... Difficult (challenging words)
AP..._ All Play (this can be any type of word)

SPECIAL NOTE: Any word preceded by a triangle (>) is designated as an All Play word. (Read the All Play section thoroughly.)

The starting picturist selects a word card from the front of the deck and has five seconds to examine the word to be played. The timer is then turned and the picturist begins sketching clues for the team. The picturist may not use verbal or physical communication to teammates during the round. Sketching and guessing continue until the word is identified or time is up. If a guess is correct, the team continues to play by rolling the die, advancing the number of squares indicated and selecting a new card and new picturist. Any number of markers may occupy the same square. THE PICTURIST POSITION ROTATES EVERY TIME A TEAM MUST SKETCH!

If a word is not identified in the time limit, the die is passed to the left. The team receiving the die begins its turn by pulling a new card from the deck, NOT a roll of the die . The word corresponding to the square in which the marker lies is the word in play. The ONLY times the die is rolled to advance the marker is when a word is identified within the one-minute time limit, or a team is first to identify the word in any All Play situation. A TEAM MUST OCCUPY A SQUARE AS LONG AS IT DOES NOT IDENTIFY THE GIVEN WORD.

ALL PLAY
In the All Play category, the word card is shown to the picturist of each team. The All Play word is sketched simultaneously by picturists to their respective teams at the start of the timer. Regardless of whose All Play it was, the first team to identify the word earns control of the die and immediately rolls and advances the number of squares indicated. This team now continues its turn with a new word. If no team identifies the word in the time allotted, the die is passed to the left. However, this team does NOT roll the die, but begins its turn by pulling a new card and sketching the word corresponding to the square they currently occupy.
SPECIAL NOTE: The above rules apply when a triangle (P) designates a word as an All Play. Remember! Any team that first identifies an All Play word immediately receives the die and rolls, then moves the indicated number of squares and draws another card.

TO WIN
A team must reach the final All Play square for the chance to win. It must be the first team to identify the word in the same manner of play as in previous All Play categories. If it is, that team wins the game. If this is not

accomplished, the die is passed to the left (in a round when a word is not identified by any team), or to the team that first identified the word. A team that reaches the final All Play square cannot win the game by winning a round controlled by another team. It must first regain control of the die to attempt a winning word. An exact roll of the die is not required to enter this square. Normal rules apply to teams not on the final All Play square.

NUMBER OF PLAYERS

The number of players per team may be uneven if an odd number of players wish to play. In the case of three players, two teams are formed. One person must act as the picturist for both teams. This person selects word cards and sketches throughout the entire game. THE PICTURIST MAY NEVER VARY. Normal game rules apply. Play is quicker and more exciting when there are less teams, and more players per team. If more than 16 people wish to play, create a fifth team or add more players to each.

YOU CAN

Draw anything related to the word, no matter how tenuous the link; Break words down into a number of syllables; Draw mail for male and blew for blue, etc.

YOU CAN'T

Use "ears" for "sounds like" or "dashes" for the number of letters in a word; Use letters or numbers; Speak to your teammates; Use sign language.

Appendix B: Changes to Pictionary Rules

For the purposes of this experiment, the pictionary rules were changed quite substantially. The board game aspects were eliminated, and some clarification was made to the guessing procedure. Briefly, these changes are as follows:

1. All material related to the board game was removed. The Pictionary game was played using only the drawing-and-guessing aspect.

2. Competitive aspects were removed. Teams were playing to beat the time limit and nothing else. Teams were not competing against each other.

3. All the "you can" and "you can't" rules apply.

4. A word was guessed when the exact word was uttered, even if it was in the context of a larger phrase or word. So "ink" was considered to have been guessed correctly when "inkwell" was used as a guess.

Appendix C: Experiment Survey and Responses

Survey

1. How was your ability to play the game was affected in the distance case? (Circle one)

Much more difficult
More difficult
No change
Easier
Much easier

2. In the distance scenario, did you experience frustration in not being able to communicate through your actions?

Yes
No

3. If you answered "yes" for question 2, did your frustration level increase or decrease as play progressed?

Increased
Decreased

4. In the distance scenario, did you and your team find ways to compensate for not being able to see each other? If so, how?

Responses to Survey

Note that the responses for question 4 were abbreviated. Drawer is always member no 1.

Subjects:
Group 1A. (1) Dave, (2) Don, (3) Cindy, (4) Youwei
Group 2A. (1) Jo, (2) Tara, (3) Meredith
Group 1B. (1) Diana, (2) Brendan, (3) Robert, (4) Tim
Group 2B. (1) Caroline Roy, (2) Leigh, (3) Manny

1. How was your ability to play the game affected in the distance case?
Group 1A. (1) More difficult, (2) MD, (3) MD, (4) No change
Group 2A. (1) No change, (2) Easier, (3) No change
Group 1B. (1) No change, (2) No change, (3) No change, (4) No change
Group 2B. (1) More difficult, (2) More difficult, (3) More difficult

2. In the distance scenario, did you experience frustration in not being able to communicate through your actions?
Group 1A. (1) No, (2) Yes, (3) No, (4) Yes
Group 2A. (1) Yes, (2) Yes, (3) No
Group 1B. (1) No, (2) No, (3) No, (4) No
Group 2B. (1) Yes, (2) Yes, (3) No

3. If you answered "yes" for Q2, did your frustration level increase or decrease as play progressed?
Group 1A. (1) -, (2) Decreased, (3) -, (4) Decreased
Group 2A. (1) Decreased, (2) Decreased, (3) -
Group 1B. ----
Group 2B. (1) Increased, (2) Increased, (3) –

4. Did your team find ways to compensate for not being able to see each other? How?
Group 1A. (1) "+ guessers were hot, - guessers were cold, x means permute/modify word" (2) "system for hot or cold, communicate if we needed an ending to the word" (3) "close... +, moved away... -, word needed to be amended... x (4) "certain signs..."
Group 2A. (1) "Agreed on meanings of symbols, learned to show when guessers were on the right track" (2) "Agreed upon signs to indicate when guesses were close", "ways for drawer to indicate which guess to focus on". (3) "Checkmark for getting closer, arrow for brainstorming, X for wrong track."
Group 1B. (1) "Auditory clues" (2) "don't need to compensate... could hear each other... important thing was to hear each other's guesses" (3) "no real compensation" (4) "verbal main mode of communication... speaking on phone helps"
Group 2B. (1) "Used indicators to lead the guessers in right area, used red pointer to signal area", (2) "indicators for right or wrong direction", (3) "indicators to lead right direction, guessed randomly and quickly"

Appendix D: Detailed Statistical Analysis

A repeated-measures analysis was performed to examine the effect of setup on within group performance. The results were insignificant, however, there was a trend indicating that groups perform better in the face-to-face setting.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TREATMEN	Sphericity Assumed	1683.062	1	1683.062	1.094	.302
TREATMEN * ORDER	Sphericity Assumed	601.678	1	601.678	.391	.536
Error(TREATMEN)	Sphericity Assumed	56940.117	37	1538.922		

Subsequently, a 2x2x2 Factorial Anova was run to potentially factor out any interesting results. There was a significant effect of day on ability, however, this result is most likely due to the limited number of subjects used in the experiment. The only implication of this result is that there exists a significant difference between the subjects that existed prior to treatment. This would potentially invalidate any results we would have determined had any been revealed.

Tests of Between-Subjects Effects

Dependent Variable: Time

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
ORDER	75.222	1	75.222	.049	.825
DAY	14812.515	1	14812.515	9.727	.003
EXPERIMENTAL SETUP	4351.660	1	4351.660	2.858	.095
ORDER * DAY	4.877	1	4.877	.003	.955
ORDER * SETUP	324.000	1	324.000	.213	.646
DAY * SETUP	1457.995	1	1457.995	.957	.331
ORDER * DAY * SETUP	154.959	1	154.959	.102	.751
Error	108121.989	71	1522.845		
Total	402854.000	79			

Looking only at the first half of the data – that is, considering only the results of each group's first setting (Face-to-face for group A's and Distance for group B's), a simple analysis could be run to look for a between group difference based on setting.

Tests of Between-Subjects Effects

Dependent Variable: WORD

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
GROUP	2572.917	1	2572.917	1.536	.223
Error	61965.750	37	1674.750		
Corrected Total	64538.667	38			