More than Face-to-Face: Empathy Effects of Video Framing

David T. Nguyen
Accenture Technology Labs
50 W San Fernando St, Ste 1200
San Jose, CA 95113
david.t.nguyen@accenture.com

John Canny
Berkeley Institute of Design
University of California, Berkeley
Berkeley, CA 94720-1774
jfc@cs.berkeley.edu

ABSTRACT
Video conferencing attempts to convey subtle cues of face-to-face interaction (F2F), but it is generally believed to be less effective than F2F. We argue that careful design based on an understanding of non-verbal communication can mitigate these differences. In this paper, we study the effects of video image framing in one-on-one meetings on empathy formation. We alter the video image by framing the display such that, in one condition, only the head is visible while, in the other condition, the entire upper body is visible. We include a F2F control case. We used two measures of dyad empathy and found a significant difference between head-only framing and both upper-body framing and F2F, but no significant difference between upper-body framing and F2F.

Based on these and earlier results, we present some design heuristics for video conferencing systems. We revisit earlier negative experimental results on video systems in the light of these new experiments. We conclude that for systems that preserve both gaze and upper-body cues, there is no evidence of deficit in communication effectiveness compared to face-to-face meetings.

Author Keywords
Video Conferencing, Empathy, Oneness

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: Miscellaneous

INTRODUCTION
Video conferencing has long been hailed as a cost-effective alternative to travel and face-to-face meetings. However, many experiments over the last several decades have found significant differences between the two experiences. Prior work has shown that non-verbal communication cues are extremely important. A recent and thorough comparison was conducted by Bos et al. [3], who found a gap between video and F2F under several measures of trust. By contrast, Nguyen and Canny found no difference under the same measures using a new system called MultiView [18]. MultiView uses a custom projection display that provides distinct spatially faithful views to each participant. It is the first system to preserve gaze cues between two groups of participants. In addition to spatial fidelity, another difference between the MultiView system and those of most prior studies (e.g. [3]) is framing. MultiView presents life-sized, upper-body images at typical meeting distances allowing MultiView to reproduce body language cues as well as gaze. We conjectured that this difference in framing should have a significant effect on communication, and, in this paper, we study that effect of framing on empathy formation.

Figure 1. The apparatus used in the experimental setup, in head-only framing condition. An echo-canceling microphone sits on the desk. Also on the desk are all the props used in the experiment: a stack of books, pens, and paperwork.

Non-verbal Communication
A natural theoretical framework for designing video conferencing systems is the field of non-verbal communication (NVC) in psychology [4, 8, 9]. The main channels of non-verbal communication are facial expression, gaze, posture, gesture, and proxemics. Video conferencing systems that frame only the head capture the

CHI 2009, April 4 - 9, 2009, Boston, Massachusetts, USA.

first two; the others require a full body or at least an upper body image (similar to F2F meetings at a table).

In live meetings, people focus most of their attention on their interlocutor's face. They also spend most of their conscious effort as "senders" on managing gaze and facial expression. These cues correlate strongly with the sender's verbal message.

People often describe in-person meetings as "face to face" as though the face were all that mattered. However, the NVC literature suggests body language cues, beyond facial cues, to be extremely important if we are to approximate in-person experiences. While facial expression and gaze are rich non-verbal channels, they are also often redundant as people are adept at controlling their facial expressions – they communicate more of what the sender intends. The other non-verbal channels convey an independent message and have been shown to carry major deception cues [8] or be instrumental in building empathy [9].

While verbal communication is relatively structured – interlocutors take turns speaking and listening – non-verbal communication is more nuanced. Two empathic people may engage in "mirroring" where proxemics and posture of one individual copy those of the other. The body may turn toward or away from the other based on liking. Arms may be open or folded based on openness and trust. Many of these gestures are directional and, just as with gaze, it is important that they not be omitted and subject to spatial distortion.

The trope of a "face-to-face meeting" may be to blame for the emphasis on head-only video in the design of video conferencing systems. The trope is quite misleading: humans rarely, if ever, have head-only encounters. Body language plays a major role with in-person encounters. With the ready availability of large-screen displays capable of presenting whole-body video, we believe it is important to understand the advantages this presentation might have.

Studying Empathy

We chose to depart from previous video conferencing studies that used prisoners' dilemma trust measures. Given the strong role that NVC plays in deception and liking, it seems likely that removing body language cues would measurably reduce trust. This is the probable cause of the differences in fragile and delayed trust between the experiments of Bos et al. without body cues [3] and Nguyen and Canny with body cues [19]. Given this prior information and the widespread use of trust studies in video conferencing, we did not feel such a study would be very illuminating. There is also a danger in relying on just one measure of effectiveness given the richness of human communication. For this reason, we sought other measures.

Empathy has been shown to affect the way people interact with each other in many different ways. For instance, empathy has been shown to have a profound role in psychotherapy. Roger's identifies it as one of the "necessary and sufficient conditions of therapeutic personality change" [20]. Empathy also plays a role in the satisfaction of relationships. Fincham et al. shows that empathy predicts positive marriage qualities [10]. Eisenberg and Miller found a relationship between empathy and both prosocial and cooperative/socially competent behavior [7]. Stephan and Finlay have shown that empathy plays a role in improving intergroup relations [22].

With prior work showing the effects of empathy on a wide range of human behavior, we decided to study whether the exclusion of body cues in video images would affect empathy. We used three measures of empathy, and two of these produced significant differences.

CONTRIBUTIONS

The main contributions of this paper are the following:

- Show that framing is an important design criterion for video conferencing systems.
- Introduce new measures of communication effectiveness based on empathy.
- Show that upper-body framing improves empathy measures and gives results not significantly different from face-to-face under several empathy measures.
- Present design suggestions for video conferencing systems based on these results and earlier studies of non-verbal communication in video conferencing.

PRIOR WORK

The importance of gaze fidelity for video conferencing has long been recognized by designers. Gaze error is a serious problem because not only are intended cues lost, but also unintended cues may be communicated: downcast eyes, sideways gaze, or gazing "over someone's head" replaces what should have been direct eye contact. These errors can result in the wrong message being communicated. The most common gaze error is downcast eyes, which communicates social deference, evasion, insincerity or boredom. The value of body language cues has been assumed in earlier works on video conferencing [3, 5, 11, 14, 15, 21, 24] but not systematically charted.

Only recently were there careful psychophysical measurements of human sensitivity to gaze errors by Chen [6]. It is remarkably asymmetric: viewers perceive almost any nonzero upward or lateral gaze errors (from 1°), but tolerate downward gaze errors of close to 10° (at 50th percentile of subjects). Fortunately, downward gaze errors arise from cameras placed above the monitor at achievable distances. Chen's paper is an important landmark because it was the first time that the distance threshold for perceived gaze error was charted. In reviewing earlier published studies, it appears that very few systems actually avoided perceptible error, even those which were described as gaze-preserving. We believe that this, in part, explains several earlier negative results. With the development of large displays, smaller cameras, and especially laptops with

integrated cameras, it is now much easier to avoid gaze error.

Several systems were designed to preserve gaze. Hydra [21] supports multi-party conferencing by providing a scaled-down camera/display surrogate that serves as a proxy for a single remote participant. Hydra correctly reproduces lateral gaze but unfortunately uses below-screen cameras, the direction in which humans are most sensitive. This results in perceptible gaze error that is unavoidable.

MAJIC [14] uses multiple cameras behind a semitransparent projection screen to accurately align the gaze of remote participants with the eyes in their video image. That paper reports on user responses to scaling and boundaries between video images in MAJIC. The scaled images studied there included head-only and upper-body framing. The upper-body framing generated higher scores to questions such as "I was able to understand what other participants did," "the conversation was natural," and "it seemed like other participants were able to communicate naturally." Head-only framing, scaled to fill the screen, gave much lower scores than upper-body framing on the same screen.

ClearBoard [15] uses a half-silvered screen to allow alignment of a camera with the remote participant's eyes. It also has an upper-body framing and the screen doubles as a shared whiteboard surface.

GAZE-2 [24] is another system developed to support gaze awareness in group settings. GAZE-2 uses an eye tracking system that selects from an array of cameras the one that the participant is looking directly at to capture a frontal facial view. Gaze direction is synthesized by placing the image on a flat virtual screen, and rotating that screen. Unfortunately, because of a perceptual phenomenon known as perspective invariance [25], the perceived gaze direction is not changed by the rotation except at very high angles resulting in gaze errors. Perspective invariance is also sometimes known as the Mona Lisa Effect [11].

MultiView [18] is a system that faithfully reproduces directional non-verbal cues, such as gaze, for several participants at each site in a group-to-group setting and shows full upper-body images. It accomplishes this by introducing a large, multiple-viewpoint display into the design of the system that give each participant their own large, unique perspective of the remote site. A trust study of MultiView was an outlier relative to earlier studies: there was no evidence of a difference between MultiView meetings and F2F, even under measures (fragile trust, delayed trust) that had previously shown such difference in desktop systems [19]. This motivated the present study.

Commercial VC systems

While it is technically trivial to send whole-body video, desktop systems emphasize "face-to-face" interaction and rarely capture video below the shoulders. Capturing a whole-body image would require a camera on top of a monitor with an extremely wide field of view. The resulting fish-eye distortion may be unpleasant for users and may need to be mitigated.

A second challenge for desktop systems is vertical gaze error. Chen recommends above-monitor camera mounts with vertical error of 5° or less [6]. At 24" monitor viewing distance, 5° is a displacement of 2". For CRT monitors with their wide boundaries and older cameras, it is virtually impossible to eliminate the error at normal viewing. The situation is much better today thanks to laptops with integrated cameras and large-screen TVs. But many studies were conducted before Chen's results and with older hardware which virtually guaranteed perceptible gaze error.

The other major conferencing market segment is conference room systems intended for group-to-group meetings. These feature a single or multiple camera(s) and large monitor(s) that capture upper-body images for several people. However, as explained by Nguyen and Canny, these systems cannot faithfully reproduce gaze and spatial gesture for more than one person per site [18]. Typical commercial designs create large lateral gaze errors (40°) for some participants.

HYPOTHESES

In this experiment, we make the following hypotheses drawing on non-verbal communication principles:

Hypothesis 1 (H1) Dyads meeting through a video conferencing system with upper-body framing will exhibit higher levels of empathy than dyads meeting with head-only framing.

Hypothesis 2 (H2) Dyads meeting face-to-face will exhibit higher levels of levels of empathy than dyads meeting with head-only framing.

While we do not have a specific mechanism to invoke, it is natural to make the following hypothesis based on general differences between video conferencing and face-to-face:

Hypothesis 3 (H3) Dyads meeting face-to-face will exhibit higher levels of empathy than dyads meeting with upper-body framing.

METHOD

System Design

In addition to F2F, we needed to design video systems with upper-body and head-only framing. For upper-body framing, we used a large-screen display with a life-sized image at a typical meeting distance of about 6'. For head-only framing, we had to choose between a life-sized, cropped image and full-screen, enlarged image. We chose the life-sized, cropped image for several reasons:

- Enlarged, head-only video performed poorly in the MAJIC study [14].
- Enlargement changes the perceived distance of the subject [14], distorting proxemic non-verbal cues.

- An enlarged image on the same screen is both enlarged and cropped. Two geometric transformations have been applied and the two effects are confounded in this case.

Full-screen face images also make it harder to avoid vertical gaze error. The eyes are near the vertical center of the head and about 18" below the top of our display. Forehead cropping would be needed to avoid gaze error.

Participants

Participants were recruited by email from a list of potential subjects maintained by our campus. They opted in by signing up via an online calendar. We scheduled 90 one-hour sessions over the course of two weeks. Each session required exactly two participants, but, due to absenteeism, up to four participants were allowed to sign up for each session. The first two participants to arrive were selected for the experiment. Participants showing up after the two required participants were compensated with a $5 show-up fee, given a Ferrero Rocher chocolate, and allowed to sign up for open slots in future sessions. For sessions where less than two participants showed, the session was canceled. Any participants showing up to a canceled session could not be considered for future sessions. Participants who participated in the experiment were compensated with $15.

Out of the 90 originally scheduled sessions, 62 were completed successfully. Data from 2 sessions were discarded because the participants' actions significantly deviated from the instructions, 25 were canceled due to absenteeism, and 1 due to technical difficulties with the video conferencing hardware. There were 124 participants who took part: 76 female (61%), 48 male (39%), 123 graduate and undergraduate students (99%) and 1 staff member (1%). These participants formed 62 pairs: 23 female-female (37%), 19 female-male (31%), 11 male-female (18%), and 9 male-male (14%) pairs.

Apparatus

Since video conferencing is a rich design space and many aspects of this space can affect communication, we give a detailed description of our system here. There were two video conferencing systems set up in separate, but adjacent rooms. Both sites are identical and illustrated in Figure 1.

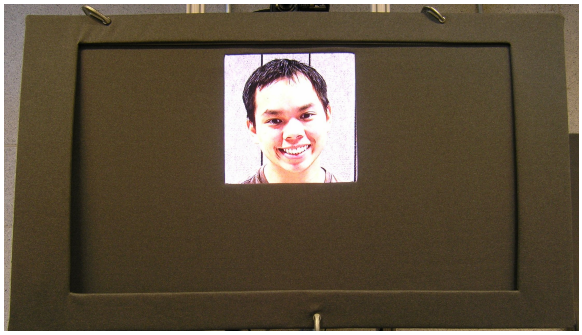
Each system consisted of a 46" 1080p high-definition liquid crystal display (Sony KDL-46XBR3) and an accompanying 1080i high definition camera (Sony HDR-SR7). The camera was mounted directly above the center of the screen. Cameras were zoomed so that the image of the remote participant was life-sized.

Audio was captured using a noise and echo-canceling microphone (ClearOne AccuMic PC). The internal speakers of the display were used for audio output.

The two systems were directly connected together. That is, the output from the camera at one site was connected directly to the display at the other site using component video and RCA audio cables. This allowed us to avoid any latency introduced by audio/video codecs and networks.

(a) Face-to-face (F2F) condition

(b) Upper-body framing (UBF) condition



(c) Head-only framing (HOF) condition

Figure 2. The three meeting conditions in the experimental design.

The participants sat about 6' from the video conferencing system in front of a large desk. In addition to the microphone, the participant also had a stack of books, three pens, consent materials, surveys, and the discussion prompt.

As with most standard video conferencing systems, there exists a displacement between the image of the remote person's eyes and the camera. In our system, there is about 8" of difference. Since the participants sit about 72" away, there is an angular disparity of about 6°. More than 90% of subjects should perceive direct eye contact at this angular disparity [6].

Treatment Conditions

The dyads met in one of three different meeting conditions (Figure 2) in a between-subjects design. Two video conferencing conditions that varied the framing of the remote partner while the third was face-to-face. The framing in the video conferencing conditions was varied by placing a mask over the display with different window sizes. Other factors, like the zoom of the camera and screen distance, remain unchanged.

Face-to-Face (F2F) The two participants of the dyad met in the same room. The seating arrangement was identical to that in the video conferencing conditions (Figure 2(a)).

Upper Body Framing (UBF) In this condition, the two participants of the dyad met in separate rooms through the video conferencing system. The mask on this video conference screen presented the viewer with a 40"x22" window of the remote scene, enough to see the entire upper body including head, torso, and hands (Figure 2(b)).

Head-Only Framing (HOF) In this condition, the two participants of a dyad met in separate rooms through the video conferencing system. The mask on this video conferencing screen presented the viewer with a 12"x12" window of the remote scene, just enough to see the head and shoulders of the remote participant (Figure 2(c)).

Measurement Instruments

In order to measure the effect on empathy, several measures were administered. Before exposure to the treatment conditions, a self-report measure known as the "Empathy Quotient" was given to control for variation in individual empathy traits [2]. After exposure to the treatment condition, we employed two different types of measures. First was a behavioral measure known as the "pen-drop experiment" that was designed to measure a person's propensity for automatic action in helping behavior [16]. Second was a self-report measure known as the "Oneness Questionnaire" which was designed to measure group cohesiveness [1]. We followed the experiment with an open-ended post interview.

Empathy Quotient

The Empathy Quotient measures individual empathy [2]. It is a Likert-type scale and includes 60 items total: 40 of which are empathy items and 20 filler/control items to relieve the participant of the relentless focus on empathy. Each item presents a statement (e.g. "When I was a child, I enjoyed cutting up worms to see what would happen.") and four choices (i.e. "strongly agree", "slightly agree", "slightly disagree", and "strong disagree"). The questionnaire asks the participant to read the statement and circle their level of agreement. The Empathy Quotient returns a score from 0 to 80.

The Empathy Quotient has been widely used to measure empathy as a personality trait. It was originally developed by Baron-Cohen and Wheelwright to classify adults with Asperger Syndrome or high functioning autism. These adults have been clinically reported to have difficulties with empathy compared to other adults. In the same study, differences in empathy were detected between sexes [2].

The Empathy Quotient was designed to measure empathy as a personality trait and should not be used to measure empathy formation toward another person. In our analysis, it was a random factor that we used as a group comparability check. Actual effects of interaction on empathy were measured using the oneness questionnaire and the pen-drop experiment.

Oneness Questionnaire

The first measure of empathy formation was the Oneness Questionnaire, which measures cohesiveness of a group and was introduced in Bailenson and Yee [1]. It is a Likert-type scale and includes ten items. Six items present a statement (e.g. "Please indicate to what extent you would use the term 'WE' to characterize you and the other person in this group by circling the appropriate number.") and a seven-point response scale with content specific endpoints (e.g. "not at all" and "extremely"). Three items are similar to the other six, but on a nine point scale. In addition, one item presents the participant with seven diagrams of two circles of varying overlap. It asks the participant to circle which set of circles best represents the dyad. The Oneness Questionnaire returns a score from 0 to 76.

The Oneness Questionnaire was used by Bailenson and Yee to perform a longitudinal study on the effects of collaborative virtual environments [1]. Elements of this oneness questionnaire include a two-item questionnaire introduced by Maner et al. in studies of helping behavior [17]. We use the Oneness Questionnaire as a self-reported attitudinal measure of empathy because of the established link between empathy and oneness. For instance, Stephan and Finlay establish the role empathy plays in improving intergroup relations as well as developing several programs based on improving empathy with the goal of intergroup relations [22].

Pen Drop Experiment

The Pen Drop Experiment measures a person's propensity for automatic action in helping behavior toward a person in distress [16]. In this experiment, a trained confederate drops pens near an unknowing participant. The original measure is whether the participant helps the trained confederate pick up the pens. In our experiment, we also measured the amount of time it took to pick up the pens when it happened.

The Pen Drop Experiment used in this study was used by van Baaren et al. to show that mimicry behavior during a conversation improves the chance of helping behavior [23]. In that study, confederates mimicked body orientation, arm position, and leg position. We note that few desktop video systems today would convey these cues. Macrae and Johnston also used the pen drop experiment to show that inhibitory cues or competing goals may eliminate priming of automatic action [16].

We use the pen drop experiment to measure empathy because of the established link between empathy and helping behavior. Eisenberg and Miller found that the formation of empathy, as measured by heart rate and facial markers, predicted prosocial behavior in children as young as second grade. Those who exhibited higher levels of empathy were associated with helping behavior that was more than the minimum amount allowable in their experiments [7].

Post Interview

The post interview follows up with participants about their experience and elicits feelings about their partners in a qualitative manner. There were no predetermined questions, but the topics covered included general impressions of the video conferencing system, their feelings on how their meeting went, and what conditions would affect whether or not they would choose to meet over video conferencing vs. face-to-face meetings. We used this information to explain some observed events during the meeting and to guide future research.

Procedure

The procedure took one hour for each session. All participants were asked, when recruited via the electronic calendaring system, to show up in the lobby of our building where they were greeted by the researcher. We immediately escorted each participant to a room to minimize the possibility of face-to-face contact between participants prior to the experiment. Roles were assigned in alternating fashion. In one session, the participant arriving first would be assigned as "the dropper" and would be trained as the pen-dropping confederate later in the experiment. In the following session, the first participant would be assigned as "the helper" and this person's pick-up behavior was observed. In each of the assigned room, the researcher presented the participant with the consent materials and went over the details of the experiment explaining that there will be a meeting with another participant through the video conferencing system, a set of surveys, and an interview. At this point, the information about the pen-drop experiment was intentionally left out as part of the deception.

Each participant was then asked to fill out the Empathy Quotient survey.

The participants were then told that they would be meeting another participant and having a 20-minute discussion. The participants were asked to discuss how they would jointly manage a $500,000 fund toward a cause of their choosing. They were prompted with twelve questions to guide them. Once the participants were comfortable with their understanding of the prompt, the participants were connected together. In the video conferencing conditions, this meant the systems were turned on and connected. In the face-to-face condition, this meant that the two separated participants were brought into the same meeting room. The participants were prompted to proceed with introductions and their discussion. Toward the end of the 20 minutes, the participants were given a 2-minute warning allowing them to wrap up and say goodbye. The conversation was ended after 20 minutes and the participants were disconnected. In the video conferencing conditions, this meant the systems were disconnected and turned off. In the face-to-face condition, the participants were separated into their respective rooms. After the discussion, participants were asked to fill out the Oneness Scale.

Upon completion of the Oneness scale, the researcher performed post interviews with both the dropper and the helper. In this post interview, the dropper was informed about the pen drop portion of the experiment and trained on how to drop the pens while carrying a large stack of books. Training was necessary to maintain consistency through the sessions. In the post interview for the helper, the researcher pointed out a stack of random books on the table and told the participant that she would need that stack of books for the next portion of the experiment. This was to reduce wonder as to why the dropper would be carrying in a stack of books when she entered.

The researcher then led the dropper – who was now carrying a large stack of books as well as three pens – to the room where the helper was seated. Before entering the video conferencing area, out of the view of the helper, the researcher pointed to the exact location as to where to drop the pens. It was marked by tape on the floor. The dropper entered the video conferencing area while the researcher stayed out of the helper's view to see if the pens were picked up and how long it took. Timing began when the pens hit the floor and timing stopped when the helper touched the pens. At that point, the researcher entered the video conferencing area. If 10 seconds had passed without the helper picking up the pens, the researcher entered the video conferencing area, picked up the pens, and sat down with the two participants.

The participants were then debriefed and filled in on the details of the experiment including the pen-drop portion. The participants were allowed to ask any questions. They were also given the option to opt-out of having the collected data used in the analysis. They were assured that we were simply measuring the effects of the video conferencing system and that all the collected data would be strictly anonymous. They were then presented with an audio-video records release form indicating how they would like the recorded materials to be handled.

Once completed, each participant was compensated with $15 as a thank you for their participation in the study and asked not to discuss the details of the study with others.

RESULTS AND ANALYSIS

Empathy Quotient

The first measure of the analysis was calculating the Empathy Quotient for each of the participants before they had a chance to meet. Because empathy is highly associated with measures of oneness and automatic action, we wanted to be careful to make sure random assignment adequately randomizes the participants along this particular character trait. Each participant's score was considered

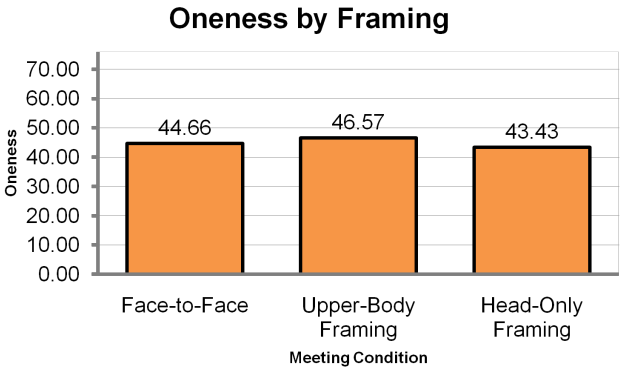


Figure 3. The self reported measure of oneness by meeting condition. Partners with head-only framing reported significantly lower scores than those with upper-body framing (p = 0.04).

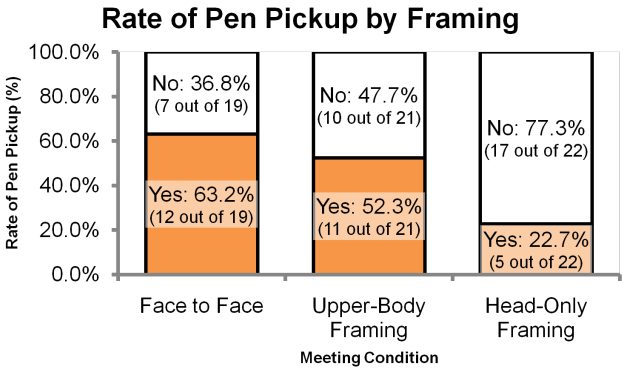


Figure 4. Pen Pickup rate. Subjects with head-only framing had significantly lower pickup rates than those with upper-body framing (p = 0.04) and those who met F2F (p = 0.01).

individually. An omnibus ANOVA failed to show any evidence that a significant difference existed between the upper-body framing condition ($M = 40.19$, $SD = 9.63$), the head-only framing condition ($M = 39.16$, $SD = 8.86$), and the face-to-face condition ($M = 40.03$, $SD = 10.46$), $F(2, 121) = 0.14$, $p = 0.87$.

Oneness Questionnaire

We continue our analysis by considering how framing affects the sense of oneness between the two participants. In our experiment, we measured oneness with a 10-item questionnaire proposed by Bailenson and Yee [1]. The reliability for this measure was good according to our data (Cronbach's alpha $= 0.858$) and no questions were excluded in the analysis.

Since the oneness scores from both the helper and the dropper were collected before the dropper was trained as the confederate, both the dropper and helper have been exposed to exactly the same treatment up until this point. Because of this, oneness scores from both participants were considered in our analysis. However, for an accurate analysis, we must take into account the possibility of partners affecting each other's score and analyze the data for dyad-level effects. We use the dyad-level approach described by Griffin and Gonzalez [13].

According to our results framing had a small but significant effect on the self-report of oneness, $Z = 1.71$, $p = 0.04$ (Figure 3). Those with upper-body framing reported an average oneness score of 46.57 ($SD = 9.24$) while those who had head-only framing reported an average of 43.43 ($SD = 9.38$). No significant difference was found between face-to-face and upper-body framing, $Z = 0.78$, $p = 0.22$. Additionally, no significant difference was found between face-to-face and head-only framing, $Z = 0.49, p = 0.31$.

Pen Drop Experiment

Next, we consider how framing affects propensity for automatic action using the pen drop experiment.

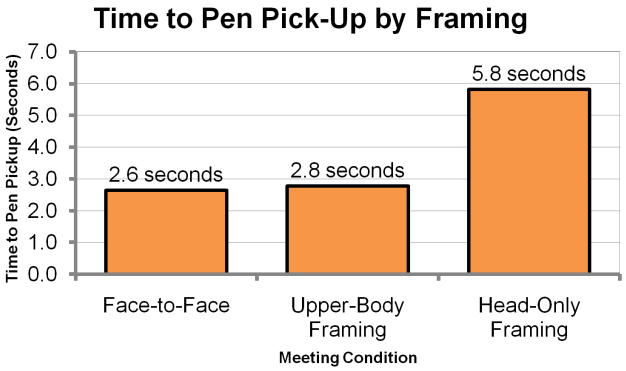


Figure 5. Pen pickup time. Subjects with head-only framing took significantly longer to pickup than those with upper-body framing (p = 0.01) and those who met F2F (p = 0.003).

First, we considered the rate of pen pickup by the helper (Figure 4). Planned Comparisons show that the meeting condition had a significant effect on whether a helper picked up their partner's pen. The rate of those in the head-only framing condition (5 out of 22) was significantly lower than the rates of those in both the upper-body framing condition (11 out of 21, $p = 0.04$, Fisher's Exact Test) and those in the face-to-face condition (12 out of 19, $p = 0.01$, Fisher's Exact Test). We failed to find a significant difference between those in the upper-body framing condition and the face-to-face condition ($p = 0.34$, Fisher's Exact Test).

Additionally, considering only when helpers picked up their partner's pens, Planned Comparisons showed a similar pattern of differences (Figure 5). The pickup time of those in the head-only framing condition (5.8 seconds) was significantly higher than the pickup times of those who met in both the upper-body framing condition (2.8 seconds, $F(1, 14) = 9.83$, $p = 0.01$) and those in the face-to-face condition (2.6 seconds, $F(1, 15) = 13.08$, $p = 0.003$). We failed to find a significant difference in pickup time between those in the upper-body framing condition and the face-to-face condition ($F(1, 21) = 0.04$, $p = 0.85$).

Finally, we used a compound statistic to compare all users by assigning users who did not pick up the pens the maximum time observed for users who did. This statistic should be more sensitive than the two above, since it combines pickup time and the action of picking up or not. However, it is clearly not normally distributed. Therefore, we used a non-parametric test. Monte-Carlo permutation tests [12] were performed using the means of synthetic pickup times as the statistic. 100 million samples of permutation distributions were taken, and p-value estimates were accurate to a relative error of 1%. Participants in the head-only framing condition produced a significantly larger value of this statistic ($p < 0.00004$) than those in the F2F condition, and significantly larger than those in the upper-body framing condition ($p < 0.0013$). There was no significant difference between those in the upper-body framing condition and the face-to-face condition ($p = 0.25$).

DISCUSSION

The pen-drop experiment produced support for both Hypothesis 1 and Hypothesis 2, that empathy was lower for groups meeting with head-only video than upper-body video or F2F. Results were significant based on the number of subjects who picked up, time to pick up among those who did, or the composite statistic which assigned the maximum time to those who did not pick up. On the other hand, there was no support for Hypothesis 3, that face-to-face produces higher empathy than upper-body video, under any measure.

Based on the Oneness Questionnaire, there was support for H1 but not H2. There was a difference in the direction of H2, but it was weaker than the significance threshold of $p = 0.05$. Since H1 is almost surely a stronger hypothesis than H2 (that is, empathy should be at least as strong in face-to-face meetings as in upper-body-framed video), we would expect to find support for H2 given a larger sample. There was no support for H3.

From this experiment, we found effect sizes from all versions of the pen drop experiment that were much larger than from the Oneness Questionnaire. That is not enough to conclude that the pen drop is a more sensitive measure of empathy. But it is interesting that the pen-drop experiment is based on an immediate (non-reflective) reaction, unlike the survey, which is based on conscious thought. Given that NVC appears to rely heavily on subconscious processes, this reinforces the importance of using measures that go beyond self-assessment. For the time being, we can say that both tests are capable of producing significant results, and so both have merit in future studies of empathy.

Both our video systems had gaze errors below the perceptible threshold [6], and so empathy should not have suffered due to gaze errors. Our results show that for video systems that present upper-body video without gaze error, there is no detectable difference under the two empathy measures when compared to face-to-face meetings. The MultiView system [18] also generated upper-body video with no gaze error and showed no difference relative to face-to-face meetings under several trust measures.

As we discussed in our review of commercial systems, off-the-shelf desktop video systems using CRT monitors and standard cameras almost surely give perceptible gaze error at normal desktop viewing distance. They also present head-only video in most cases. Systems such as MAJIC [14] and ClearBoard [15] eliminate gaze errors, but earlier studies of the effectiveness of video conferencing have used off-the-shelf hardware. Framing is usually head-only or head-and-shoulders, and gaze errors appear to have been likely.

DESIGN GUIDELINES FOR VIDEO SYSTEMS

In order to preserve gaze and upper-body cues, video conference systems need to be carefully designed and set up. It is relatively easy to preserve both with new hardware, but also easy to break them. We consider three cases: one-on-one desktop conferencing, one-on-one large-screen conferencing, and many-to-many large screen.

Desktop/Laptop design

Desktop conferencing has been given a boost with the introduction of laptops with integrated cameras. These cameras sit very close (1/2" or less) to the screen, and above it. Apart from convenience, they dramatically reduce the gaze disparity compared to set-top webcams. Ideal desktop/laptop monitor viewing distance is less than 2'. According to Chen, the recommended gaze error is 5° or less [6], which translates to 1" of screen height for every foot of viewing distance. Therefore, the maximum distance between camera and image of participant's eyes should be 2" at 2'. An integrated camera at 0.5" above the image leaves 1.5" of image space to accommodate the space between the participant's eyes and the top of their head.

However, if one shows a large head-only image on the monitor, gaze error is likely. The eyes in a human face sit close to the vertical halfway-point (e.g. head height 9", eyes to top-of-head 4.5"). However, if the image is scaled and it fills the screen vertically, the eyes will fall near the vertical middle of the screen. This is much further than the 1.5" allowed from the top of the screen. A solution is to scale the image down, and place it at the top center of the screen.

If one instead shows an upper-body image on the monitor, things work much better. A view of 27" of the subject's upper body is six times the eye-to-top-of-head distance. If the image is scaled to fill the vertical screen space of a 6" high monitor (with the head just touching to top), the eyes will be just 1" below the top of the screen. Even for a large monitor (9" high) and similar scaling, one is still at the recommended gaze error distance. In fact, there is some margin for error with upper-body images. The 90% sensitivity limit in is actually 8° [6], so most users will not detect gaze error until the desktop eye-camera distance is over 3". As long as an upper-body image is placed to fill the vertical distance with the head *somewhere* near the top of the screen, there should be no gaze error effects.

A difficulty with current laptop cameras is that they are designed to frame *head-only images* at normal distances.

Figure 6. Laptop makers should choose camera optics to capture an upper body view (right) rather than just the face (left).

So one cannot obtain an upper-body image unless your remote correspondent moves 2 to 3 times the normal viewing distance from their laptop. Since the image they see is already well below life-size, this will probably be unacceptable to most users.

The solution (a plea to laptop makers): choose camera optics to frame upper-body images at normal viewing distance (Figure 6). Users have only to maximize the window to produce an image that is both gaze- and body-language preserving. Some fisheye distortion is likely, but the overall conferencing experience should dramatically improve.

TV Set-top design

Viewing distances for large screen TVs make it easier to design effective interactions. At an 8'-12' viewing distance, the recommended eye-camera distance is 8"-12". An off-the-shelf camera placed flush on top of the TV will still leave several inches of screen space to accommodate the eye-head distance. A full screen upper-body image should give close to the recommended eye-camera distance for 5° error, and certainly within the 90% allowable angle error of 8°. The camera lens can easily be chosen to frame the correspondent's upper body (or the whole body which may be more appropriate) at 8'-12' viewing distance.

Group Conferences

In many cases, users want a group-to-group conference. Certainly, in the workplace, and family-to-family conferences are natural in the home. Perspective Invariance [25] (aka Mona Lisa Effect [11]) makes this quite challenging. It leads to large gaze errors for some participants, and it is an intrinsic problem with conventional displays. MultiView displays are required to restore gaze, which have been available only in the lab. However, suitable displays are currently under development, such as the Sharp 'Triple Directional Viewpoint LCD' which provide three independent displays in three different viewing directions. Group-to-group conferencing may soon have the benefit of gaze and body language fidelity.

CONCLUSION

This paper studied the effect of framing in video conferencing. Drawing from the non-verbal communication field, we hypothesized that removal of body language cues due to head-only framing would significantly impact communication compared to upper-body framing or face-to-face meetings. We evaluated the interaction of dyads meeting through different media (head-only video, upper-body video, and face-to-face) under two measures of empathy. We found significant differences under both measures and support for our two primary hypotheses. A third hypothesis that dyads meeting face-to-face would share more empathy than dyads meeting with upper-body video was not supported.

Using these findings and earlier results on gaze-preserving video systems, we presented guidelines for conferencing systems that preserve both gaze and body language cues.

We observed that the gaze requirements [6] are difficult to meet for older hardware (i.e. CRT monitors with set-top cameras) and that previous studies of video conferencing effectiveness appear to be based on systems that did not preserve the experience of gaze, even if intended to.

Furthermore, aside from [14] the effects of body language cues have not been considered, and desktop systems normally eliminate them. This paper shows that their effect is significant, and that there are no evident differences between F2F and video systems which preserve both gaze and body language cues. The MultiView study [19] produced similar findings: no difference between F2F and gaze- and body-language preserving video under several measures of trust, and between groups of participants.

Non-verbal communication has proved to be a powerful framework for understanding the video medium. Systems that are careful to preserve important non-verbal cues provide F2F-like experiences, while those that do not are measurably inferior to F2F.

These findings call into question the "conventional wisdom" about video conferencing – that it is a poor alternative to face-to-face meetings. A significant part of the published evidence for this appears to derive from difficulties with older hardware and does not apply to current and future systems that are well-designed. More work remains to be done – especially studies of a wider variety of tasks and with richer metrics for interaction. It would be premature to declare video conferencing a universal substitute for face-to-face. On the other hand, it may be much better for suitable tasks (e.g. negotiation, trust building, socializing) than had been believed.

We hope that this paper encourages a fresh look at video conferencing as an alternative to in-person meetings. Especially given the economic and environmental cost of travel, all reasonable alternatives deserve consideration.

ACKNOWLEDGMENTS

We thank the Experimental Social Science Laboratory (XLab) at the University of California, Berkeley for their support in recruiting and managing participants in the user study, the members of the Berkeley Institute of Design (BiD) and Accenture Technology Labs for their support in designing and piloting the experimental study as well as authoring of this paper, and the anonymous reviewers for their constructive feedback.

REFERENCES

1. J. N. Bailenson and N. Yee. A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoper. Virtual Environ.*, 15(6):699–716, 2006.
2. S. Baron-Cohen and S. Wheelwright. The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2):163–175, April 2004.
3. N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright. Effects of four computer-mediated communications channels on trust development. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 135–140, New York, NY, USA, 2002. ACM Press.
4. J. K. Burgoon. *Nonverbal Communication: The Unspoken Dialogue*. Harpercollins College Div.
5. W. A. S. Buxton, A. J. Sellen, and M. C. Sheasby. Interfaces for multiparty videoconferencing. In K. Finn, A. Sellen, and S. Wilber, editors, *Video Mediated Communication*, pages 385–400. Erlbaum, Hillsdale, N.J., 1997.
6. M. Chen. Leveraging the asymmetric sensitivity of eye contact for videoconference. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 49–56, New York, NY, USA, 2002. ACM Press.
7. N. Eisenberg and P. A. Miller. The relation of empathy to prosocial and related behaviors. *Psychological bulletin*, 101(1):91–119, January 1987.
8. P. Ekman. *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised and Updated Edition)*. W. W. Norton & Company, 2 rev sub edition, September 2001.
9. S. Feldstein and A. W. Siegman, editors. *Nonverbal Behavior and Communication*. Lawrence Erlbaum Associates, 1987.
10. F. D. Fincham, G. F. Paleari, and C. Regalia. Forgiveness in marriage: The role of relationship quality, attributions, and empathy. *Personal Relationships*, pages 27–37, 2002.
11. J. Gemmell, K. Toyama, L. C. Zitnick, T. Kang, and S. Seitz. Gaze awareness for video-conferencing: a software approach. *Multimedia, IEEE*, 7(7):26–35, October-December 2000.
12. P. I. Good. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. Springer, December 2004.
13. D. Griffin and R. Gonzalez. Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*, 118(3):430–439, November 1995.
14. Y. Ichikawa, K. I. Okada, G. Jeong, S. Tanaka, and Y. Matsushita. Majic videoconferencing system: experiments, evaluation and improvement. In *ECSCW'95: Proceedings of the fourth conference on European Conference on Computer-Supported Cooperative Work*, pages 279–292, Norwell, MA, USA, 1995. Kluwer Academic Publishers.
15. H. Ishii, M. Kobayashi, and J. Grudin. Integration of interpersonal space and shared workspace: Clearboard design and experiments. *ACM Trans. Inf. Syst.*, 11(4):349–375, 1993.
16. N. C. Macrae and L. Johnston. Help, i need somebody: Automatic action and inaction. *Social Cognition*, 16(4):400–417, 1998.
17. J. K. Maner, C. L. Luce, S. L. Neuberg, R. B. Cialdini, S. Brown, and B. J. Sagarin. The effects of perspective taking on motivations for helping: Still no evidence for altruism. *Pers Soc Psychol Bull*, 28(11):1601–1610, November 2002.
18. D. Nguyen and J. Canny. Multiview: spatially faithful group video conferencing. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 799–808, New York, NY, USA, 2005. ACM Press.
19. D. T. Nguyen and J. Canny. Multiview: improving trust in group video conferencing through spatial faithfulness. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1465–1474, New York, NY, USA, 2007. ACM.
20. C. R. Rogers. The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting and Clinical Psychology*, 60(6):827–837, December 1992.
21. A. J. Sellen. Speech patterns in video-mediated conversations. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 49–59, New York, NY, USA, 1992. ACM Press.
22. W. G. Stephan and K. Finlay. The role of empathy in improving intergroup relations. *Journal of Social Issues*, 55(4):729–743, 1999.
23. R. B. Van Baaren, R. W. Holland, K. Kawakami, and A. van Knippenberg. Mimicry and prosocial behavior. *Psychological Science*, 15(1):71–74, 2004.
24. R. Vertegaal, G. van der Veer, and H. Vonsfect. Effects of gaze on multiparty mediated communication. In *Proceedings of Graphics Interface*, pages 95–102, Montreal, Canada, 2000. Human-Computer Communications Society, Morgan Kaufmann Publishers.
25. D. Vishwanath, A. R. Girshick, and M. S. Banks. Why pictures look right when viewed from the wrong place. *Nature Neuroscience*, 8(10):1401–1410, September 2005.