Out of Sight, …
How Asymmetry in Video-Conferencing Affects Social Interaction

Camille Sallaberry
c.sallaberry@utwente.nl
University of Twente
Enschede, Netherlands

Gwenn Englebienne
University of Twente
Enschede, Netherland
g.englebienne@utwente.nl

Vanessa Evers
University of Twente
Enschede, Netherlands

Jan van Erp
University of Twente, TNO
Enschede, Netherlands

ABSTRACT

As social-mediated interaction is becoming increasingly important and multi-modal, even expanding into virtual reality and physical telepresence with robotic avatars, new challenges emerge. For instance, video calls have become the norm and it is increasingly common that people experience a form of asymmetry, such as not being heard or seen by their communication partners online due to connection issues. Previous research has not yet extensively explored the effect on social interaction. In this study, 61 Dyads, *i.e.* 122 adults, played a quiz-like game using a video-conferencing platform and evaluated the quality of their social interaction by measuring five sub-scales of social presence. The Dyads had either symmetrical access to social cues (both only audio, or both audio and video) or asymmetrical access (one partner receiving only audio, the other audio and video). Our results showed that in the case of asymmetrical access, the party receiving more modalities, *i.e.* audio and video from the other, felt significantly less connected than their partner. We discuss these results in relation to the Media Richness Theory (MRT) and the Hyperpersonal Model: in asymmetry, more modalities or cues will not necessarily increase feeling socially connected, in opposition to what was predicted by MRT. We hypothesize that participants sending fewer cues compensate by increasing the richness of their expressions and that the interaction shifts towards an equivalent richness for both participants.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**.

KEYWORDS

"Mediated Communication; Social Interaction; Social Computing; Asymmetry; Video-Conference; Social Presence; Social Signals"

ACM Reference Format:
Camille Sallaberry, Gwenn Englebienne, Vanessa Evers, and Jan van Erp. 2023. Out of Sight, … How Asymmetry in Video-Conferencing Affects Social Interaction. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '23), October 09–13, 2023, Paris, France.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3577190.3614168

This work is licensed under a Creative Commons Attribution-NoDerivs International 4.0 License.

ICMI '23, October 09–13, 2023, Paris, France
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0055-2/23/10.
https://doi.org/10.1145/3577190.3614168

1 INTRODUCTION

Humans are inherently social. In face-to-face interaction, we use different social modalities like verbal communication, gestures or touch to communicate with others and express social information. Those communications can vary from a simple greeting to establishing more complex constructs such as trust. Social cues, as part of interpersonal communication, play a key role in defining and regulating our interactions and relationships with the people around us and are used to structure communication and recognize people's emotions [14]. During social exchange, the receivers interpret the social cues, taking into account different parameters such as the context of the communication or their own emotions, and express their social message in return.

In mediated interaction, accurate exchange of information is heavily dependent on the medium (*i.e.* the more modalities or cues, the more richness). With the Media Richness Theory (MRT), Daft and Lengel [6] predict a lower social connection in less rich media environments. In contrast to MRT, the formulation of the Hyperpersonal Model [23] states that people interacting in an online environment connect easier than in face-to-face interaction, even if they have access to fewer social cues and less richness of information. We test the predictions of both theories using symmetric (people access the same modalities and cues) and asymmetric (people access a different type, quality and/or number of social cues or modalities) mediated communications.

1.1 Cues and social information

Cues do not necessarily transmit the same information. App et al. [1] demonstrated that different emotions are associated with different cues and with different modalities. The preferential use of those modalities is probably due to the social functions associated with each emotion. For example, fundamental emotions such as anger, fear or happiness are displayed with a combination of most modalities, *i.e.* facial expression, touch and body movement, while complex emotions like trust, love or affection are often primarily expressed with touch [8], and pride through body posture. Modalities are also not used to the same degree. Facial expressions are often considered one of the most prominent non-verbal cues. Sharan et al. [18] describe cues used in mediated social communication (MSC) and rank

them by relevance to social interaction. Gaze comes first, followed by a combination of facial expression and prosody. Gaze is crucial for sharing attention and engagement during interaction [17], while facial expression and prosody are both important channels to display emotion. On the one hand, social modalities are complementary to each other and can be used in combination [1, 11]. For example, facial expressions are often used as confirmation and amplification of verbal information. On the other hand, a discrepancy of information between cues makes interaction ambiguous and more difficult to interpret [7].

1.2 Social richness in mediated communication

Past research demonstrated that having more media richness, including the number of available cues and modalities for communication, allows for a richer exchange of information (*i.e.* Media Richness Theory) [6]. In mediated communication, media richness varies with, and is typically limited by, the hardware. Social modalities, albeit in limited form, still demonstrate social effects during mediated interaction [2, 10]. For example, O'Malley et al. [15] compared face-to-face communication with audio-only and audio-video mediated communication. They show that audio-video communication tends to conserve the potential for non-verbal understanding, and can even lead to shorter response time between participants than in face-to-face communication. However, visual cues such as eye-gaze tend to be less effective than in face-to-face, reducing the effectiveness of social communication. With only audio, participants tend to compensate by verbalising more to allow a better understanding and acknowledgment. In remote haptic, Haans et al. [10] showed that the "Midas touch" effect (i.e. the ability to entice people to do something via touching them) was still present in mediated touch by comparing the proneness of helping people in no-touch or touch situations.

In 1996, Walther [23] proposed the Hyperpersonal model about people's self-disclosure and compensation behavior in mediated communication. They show that people in online environments tend to feel more affiliated with others and self-disclose more about their own life compared to face-to-face. In contrast, Sherman et al. [19] investigated different mediated communication effects on social bonding between friends and showed that the level of bonding and connectedness tends to go down with the reduction of non-verbal modalities and cues. These results confirmed the findings of Daft and Lengel [6] and the Media Richness Theory but, interestingly, the findings were in opposition to multiple previous studies, including the Hyperpersonal model [23]. Sherman et al. showed that the level of bonding and connectedness tends to go down when access to non-verbal modalities and cues is reduced. However, their findings concerned existing friends and not strangers. Thus, those results may have been impacted by the notion that people in existing relationships tend to trust and know each other more.

1.3 Asymmetry & mediated interaction

Asymmetry in media has gained attention in recent years. In this paper, asymmetry will be defined as *a specific condition that can happen during mediated communication when people interacting have access to a different type, quality and/or number of social cues (*e.g. *voice, facial expression, touch, body posture).* In 2008, Voida et al. discussed the presence of asymmetry in the media space and identified six forms of asymmetry: asymmetry of media, of fidelity, of participation, of engagement, of benefit and asymmetry of place [22]. They observed that in the last twenty years, most research focused on achieving symmetrical reciprocity and aimed for the mitigation of asymmetry through technical and social methods. An example of a social method is making asymmetric interaction socially inappropriate (*e.g.* seeing someone on video while yours is disconnected being considered rude). However, they also argued that asymmetry, like symmetry, should be considered as an asset instead of being rejected.

Recently, an increasing number of studies in mediated collaboration have focused on asymmetric systems, especially in the field of virtual reality (VR), mixed reality and telepresence [16, 20]. However, most of this research focuses on collaboration and task performance. For example, Kolkmeier et al. [12] introduce a mixed reality toolkit for mediated collaboration designed to allow two people at a distance to work together with one person in VR and the other in augmented reality (AR). In 2019, Grandi et al. [9] also investigated collaboration performance between people in symmetric or asymmetric VR and AR environments. What they discovered was that people in VR-AR asymmetry performed better than those in AR symmetry and slightly worse than in VR symmetry. Their results also highlighted the fact that, in their design, task collaboration was not affected by the type of interaction and the difference in visualization (AR or VR). Despite the importance of social communication, few studies measured the socialness of the interaction when investigating asymmetrical mediated interaction, and focus on the task-related aspects of the interaction [3, 12].

As access to VR, mediated haptic and telepresence using robotic avatars develops, people will use systems with a high level of social richness, making it possible to even surpass face-to-face interaction in terms of access to information, for example by integrating face recognition or physiological data. This highlights the importance of investigating the role of asymmetry in the social context and its possible effects on the interaction. Van Erp et al. [21] presented a new system for avatar-mediated interaction, using a humanoid robot. The robot can express multiple social cues such as eye gaze, facial expression, gesture and touch. Intended for remote collaboration and social interaction, it is important to note that this system introduces several asymmetries in human-mediated interaction. While the robot controller sees the real face of the remote recipient via camera, the remote recipient only sees an abstract representation of the facial expression of the controller on the robot. Similarly, while the remote recipient will be touching, the controller will receive sensations from haptic wearable devices. In this system, people inherently have access to a different richness of information and asymmetry of media, fidelity, engagement or even place [22].

Further work is needed to investigate asymmetry in remote social interaction. This paper evaluates the effect of asymmetrical access to social modalities in a mediated dyadic interaction, on each party's perception of the social presence of the other. In other words: "How does asymmetrical access to social cues in video-conferencing affect social interaction?" In symmetrical interaction, we know that more social media richness leads to better information transfer [6]. Thus, we expect that when more modalities are available, more

information will be obtained and the partner will be perceived as more present. (*i.e. During a social-mediated interaction, people accessing more modalities will rate social presence higher than with fewer modalities (H1).*)

However, according to the hyperpersonal model [23], people in mediated interactions tend to compensate for missing modalities and may feel even more connected than when interacting face-to-face. This raises the question: in which asymmetry condition of a mediated interaction (*i.e.* when receiving more cues or when sending more cues) does one feel more connected to the other. Our assumption is that people sending fewer cues will compensate more, so that the person receiving fewer modalities than their partner will rate social presence in the interaction higher (H2).

2 METHOD

Due to COVID restrictions at the time, this experiment took place online with participants interacting with each other from their respective homes. We recruited a total of 122 university students (55 female, 57 male). Most dyads were not strangers. The experiment was reviewed and approved by the Internal Review Board of the University of Twente (RP 2020-127).

2.1 Experimental design

We used access to audio and video to create the three communication groups and four communication conditions. Two groups, of 14 dyads each, had symmetrical communication — each partner in condition "receiving and sending only *audio*" ("aa"), or each partner receiving and sending *both* audio and video (condition "bb") — and one group, of 28 dyads, had asymmetrical communication: one partner received both video and audio and sent audio only (condition "ba"), the other partner received audio only and sent both audio and video (condition "ab").

We had five dependent variables, i.e. the five sub-scales of Biocca's questionnaire on social presence [4]: Loneliness (LO), Mutual Awareness (MA), Emotional Contagion (EC), Mutual Understanding (MU) and Behavioral Engagement (BE). We reduced the number of questions in Biocca's questionnaire from the original 34 to 26. All items were rated on a 5 points Likert scale. Analysis was performed on the individual level.

2.2 Procedure

Each participant was randomly paired with another to form a dyad. Dyads were randomly assigned to one of three communication groups. Participants either received course credits or €5 for their participation. At the start of the session, both participants were invited by the researcher on Zoom®, a video-conferencing software from home to interact. After some explanation, they were both asked to follow the condition of the experiment (with/without video) and start the task. The dyad played a picture quiz in which both communication partners saw different pictures of a city and had to describe each other pictures for both to guess the name of the city and earn points. Additional pictures were displayed upon request at a scoring cost. We designed the quiz using the Qualtrics® survey platform. At the end of the session, the participants filled out an online survey using a provided link. At the end of the experiment, each pair was debriefed. On average, the experiment took 45 minutes to finish for a maximum of 30 minutes spent on the task.

3 RESULTS

Out of the 61 dyads (122 participants), we removed five, e.g. because they did not follow the instructions, resulting in 28 participants in each of the four conditions. We evaluated the validity of the social presence sub-scales with the Cronbach's alpha as the original survey was shortened, see table 1. Except for MA being somewhat low (a > 0.6), most alpha values of the sub-scales are good (a > 0.7). This means that the survey, albeit shortened, is still valid.

LO	MA	EC	MU	BE
0.91	0.62	0.74	0.73	0.76

Table 1: Cronbach's alpha for the five sub-scales of social presence

Because the data did not follow the normal distribution, we conducted the Kruskal-Wallis test for both hypotheses (H1 and H2) and the Dune post-hoc analysis test for H1.

3.0.1 Hypothesis 1. For the first hypothesis, we compared participants in dyads with access either to audio only (aa) with two modalities (both can only send audio), asymmetry (ab/ba) with three modalities (calculated as the sum of sending and receiving) and audio/video (bb) with four modalities (both send and receive audio and video).

K.W.	**L**	**MA**	**EC**	**MU**	**BE**
p-value	0.178	0.0632	0.00775	0.458	0.224

Table 2: Kruskal-Wallis test's p-values for each social presence's sub-scales depending on the number of modalities

We found a significant effect on one sub-scale: Emotional Contagion (p < .01) and a trend on Mutual Awareness (p < .07), see table 2. For these two sub-scales, we did a post-hoc analysis using Dunn's method to test the sub-scales' sensitivity and compare the different groups related to our first hypothesis.

To test the sensitivity of the sub-scales, we compared the condition with audio only (aa) to the group with both audio and video (bb) to see if we can reproduce the common finding that more cues are preferred over fewer cues. The condition with both audio and video showed favorable scores on emotional contagion (p < .006, see Table 2) and a trend on mutual awareness (p < .06). We compared the asymmetry condition (ab & ba) to each symmetric condition (aa & bb). Post-hoc test showed a difference between asymmetry (ab & ba) and both audio & video for EC (p < .04). However, this effect is no longer significant after adjusting p for multiple testing. None of the other effects reached significance.

3.0.2 Hypothesis 2. To test H2 "difference between giving and receiving", we compared the two asymmetric conditions, *i.e.* giving more (ab) and receiving more (ba), with a Kruskal-Wallis test. Results showed a significant difference for loneliness only (Table 3)

K.W.	LO	MA	EC	MU	BE
p-value	.017	.101	.298	.940	.721

Table 3: Kruskal-Wallis test's p-values for each social presence's sub-scales between asymmetric conditions ab and ba

where participants sending more cues felt less lonely than the ones receiving more: median ab = 3.5, median ba = 3.0; mean ab = 3.39±.91, mean ba = 2.62±1.2.

4 DISCUSSION

Our findings show 1) that richer media (here audio combined with video) results in higher scores for Emotional Contagion and Mutual Awareness (trend) than less rich media (audio only), while asymmetry's score tends to be in between audio and audio-video, but with no significant difference with either audio or audio-video and 2) that people sending more modalities score higher on the Loneliness sub-scale (feel more connected) than people who receive more modalities but send less than their partner.

The first result partially confirms the first hypothesis for the Emotional Contagion and Mutual Awareness sub-scales and is partially in line with MRT. However, no other sub-scales resulted in significant differences between the groups, and there were no significant differences between the symmetric and asymmetric groups. Moreover, Mutual Awareness was just a trend. In relation to MRT, this finding may indicate that symmetric and asymmetric dyads follow the prediction of MRT, i.e. the more modalities or cues accessible, the more connected people may feel.

The second result confirms the second hypothesis, but only for the loneliness sub-scale. It shows that, in asymmetric interaction, participants that send more cues than they receive (ab) rate the interaction as feeling more present/connected than participants that receive more cues than they send. This finding indicates that sending and receiving social cues may differently affect partners in a dyad, and that sending cues contributes more to feeling connected than receiving them. Another explanation could be that the participants receiving audio and video, make an effort in getting eye contact with their partner but get no acknowledgement from them. This subconsciously brings the participants with a video feed to be negatively impacted by the 'disconnected' video" on the other side. Finally, it is also possible that people receiving audio and video may feel uneasy (feeling of being rude) to send less than receive. It may be seen as violating existing social norms. This finding is partially in opposition to the predictions of MRT as MRT only considers symmetric access and was not formulated to differentiate between sending and receiving cues. This also contradicts the Hyperpersonal model as the person sending more information, i.e. revealing more with less control over information, feels more connected to the other.

From previous research on asymmetric collaboration and attention in mediated interaction [18, 22], we expected significant effects on multiple social presence sub-scales. It is not yet well-studied how people's facial expressions change when people consciously know that they are seen but don't get immediate feedback from seeing their conversational partner's reaction. Previous research showed that people's facial expressions are different depending on their relationship with others or when they do not know they are observed [5, 13]. We can expect that our social signals may be affected when we feel we are virtually "behind a one-way mirror". Our task was also sufficiently "game-like" that people would tend to engage no matter what, thus suppressing what effects video might have had.

4.1 Limitations

Our study design could have been more robust with respect to the following aspects. 1) we used a condensed version of Biocca's social presence questionnaire to limit participant strain, fatigue and loss of interest (which were very relevant during Covid-19 restrictions), but this may have affected the quality of the measurement. 2) Participants needed to focus on a visual task, which may have limited their attention to the video of the other participant. 3) The study was performed during lock-down, on each individual participant's equipment, potentially leading to uneven audio and video quality across participants.

4.2 Conclusion and Future work

We conclude that there are subtle and as-yet not-well-understood but important and complex effects of asymmetry in mediated dyadic communication. The informal rule of thumb to avoid asymmetry (for example, by turning off one's camera when the other party's camera is not working), will not be applicable in advanced telepresence systems currently under development. Research on asymmetrical cues in social interaction beyond vision and audition and in relation to the Media Richness Theory and the Hyper-personal model is therefore timely and relevant.

Future research will need to consider other measures such as cognitive load and user experience to investigate the effects of asymmetry further: we didn't take into account possible compensating behaviors from participants and the effect of social norms, *i.e.* people forced to send only audio might feel uneasy. Those include also: the effects of asymmetry on task performance; the importance of context, e.g. formal vs. informal; whether collaborative and competitive interactions are affected differently; the effects of prior acquaintance; how these effects carry over to three-way communication and beyond; how power-dynamics are affected; etc. To conclude, this study highlights the importance of understanding the effects of asymmetry, as more advanced communication technologies and "hybrid" telepresence systems, including AR, VR, and robotic avatars will make asymmetric communication ever more common.

ACKNOWLEDGMENTS

We are grateful to Dasha Kolesnyk for her invaluable help in running the experiments described in this work. We also thank the volunteers who participated, and the reviewers whose comments have improved the manuscript.

REFERENCES

[1] B. App, D.N. McIntosh, C.L. Reed, and M.J. Hertenstein. 2011. Nonverbal Channel Use in Communication of Emotion: How May Depend on Why. *Emotion* 11, 3 (2011), 603–617. https://doi.org/10.1037/a0023164 cited By 78.

[2] J. N. Bailenson, N. Yee, S. Brave, D. Merget, and D. Koslow. 2007. Virtual interpersonal touch: Expressing and recognizing emotions through haptic devices.

Human-Computer Interaction 22, 3 (2007), 325–353. www.scopus.com Cited By :150.

[3] C. Bevan and D. Stanton Fraser. 2015. Shaking Hands and Cooperation in Tele-present Human-Robot Negotiation. In *ACM/IEEE International Conference on Human-Robot Interaction*, Vol. 2015-March. 247–254. www.scopus.com Cited By :36.

[4] Frank Biocca and Chad Harms. 2002. Defining and measuring social presence: Contribution to the networked minds theory and measure. *Proceedings of PRESENCE* 2002 (2002), 1–36.

[5] Pamela M. Cole. 1986. Children's Spontaneous Control of Facial Expression. *Child Development* 57, 6 (1986), 1309–1321. http://www.jstor.org/stable/1130411

[6] Richard Daft and Robert Lengel. 1986. Organizational Information Requirements, Media Richness and Structural Design. *Management Science* 32 (05 1986), 554–571. https://doi.org/10.1287/mnsc.32.5.554

[7] B. de Gelder, A. W. de Borst, and R. Watson. 2015. The perception of emotion in body expressions. *Wiley Interdisciplinary Reviews: Cognitive Science* 6, 2 (2015), 149–158. www.scopus.com Cited By :105.

[8] D.-M. Ellingsen, J. Wessberg, O. Chelnokova, H. Olausson, B. Laeng, and S. Leknes. 2014. In touch with your emotions: Oxytocin and touch change social impressions while others' facial expressions can alter touch. *Psychoneuroendocrinology* 39, 1 (2014), 11–20. https://doi.org/10.1016/j.psyneuen.2013.09.017 cited By 56.

[9] J. G. Grandi, H. G. Debarba, and A. Maciel. 2019. Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*. 127–135. www.scopus.com Cited By :39.

[10] A. Haans, R. de Bruijn, and W. A. IJsselsteijn. 2014. A Virtual Midas Touch? Touch, Compliance, and Confederate Bias in Mediated Communication. *Journal of Nonverbal Behavior* 38, 3 (2014), 301–311. www.scopus.com Cited By :36.

[11] Mark L Knapp, Judith A Hall, and Terrence G Horgan. 2013. *Nonverbal communication in human interaction*. Cengage Learning.

[12] J. Kolkmeier, D. Reidsma, E. Harmsen, M. Theune, S. Giesselink, and D. Heylen. 2018. With a little help from a holographic friend: The OpenIMPRESS mixed reality telepresence toolkit for remote collaboration systems. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. www.scopus.com Cited By :17.

[13] Robert E Kraut. 1982. Social presence, facial feedback, and emotion. *Journal of Personality and Social Psychology* 42, 5 (1982), 853.

[14] C. Marinetti, P. Moore, P. Lucas, and B. Parkinson. 2011. Emotions in social interactions: Unfolding emotional experience. *Cognitive Technologies* 9783642151835 (2011), 31–46. https://doi.org/10.1007/978-3-642-15184-2_3 cited By 12.

[15] C. O'Malley, S. Langton, A. Anderson, G. Doherty-Sneddon, and V. Bruce. 1996. Comparison of face-to-face and video-mediated interaction. *Interacting with Computers* 8, 2 (1996), 177–192. www.scopus.com Cited By :104.

[16] R. A. Ruddle, J. C. D. Savage, and D. M. Jones. 2002. Symmetric and Asymmetric Action Integration During Cooperative Object Manipulation in Virtual Environments. *ACM Transactions on Computer-Human Interaction* 9, 4 (2002), 285–308. www.scopus.com Cited By :68.

[17] A. Senju and T. Hasegawa. 2005. Direct gaze captures visuospatial attention. *Visual Cognition* 12, 1 (2005), 127–144. www.scopus.com Cited By :206.

[18] N. N. Sharan, A. Toet, T. Mioch, O. Niamut, and J. B. F. van Erp. 2022. *The Relative Importance of Social Cues in Immersive Mediated Communication*. Lecture Notes in Networks and Systems, Vol. 319. 491–498 pages. www.scopus.com

[19] L. E. Sherman, M. Michikyan, and P. M. Greenfield. 2013. The effects of text, audio, video, and in-person communication on bonding between friends. *Cyberpsychology* 7, 2 (2013). www.scopus.com Cited By :90.

[20] A. Steed, W. Steptoe, W. Oyekoya, F. Pece, T. Weyrich, J. Kautz, D. Friedman, A. Peer, M. Solazzi, F. Tecchia, M. Bergamasco, and M. Slater. 2012. Beaming: An asymmetric telepresence system. *IEEE Computer Graphics and Applications* 32, 6 (2012), 10–17. www.scopus.com Cited By :39.

[21] Jan BF Van Erp, Camille Sallaberry, Christiaan Brekelmans, Douwe Dresscher, Frank Ter Haar, Gwenn Englebienne, Jeanine Van Bruggen, Joachim De Greeff, Leonor Fermoselle Silva Pereira, Alexander Toet, et al. 2022. What Comes After Telepresence? Embodiment, Social Presence and Transporting One's Functional and Social Self. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2067–2072.

[22] A. Voida, S. Voida, S. Greenberg, and H. A. He. 2008. Asymmetry in media spaces. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. 313–322. www.scopus.com Cited By :40.

[23] Joseph B. Walther. 1996. Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction. *Communication Research* 23, 1 (1996), 3–43. https://doi.org/10.1177/009365096023001001 arXiv:https://doi.org/10.1177/009365096023001001

accepted 21 July 2023