Beyond Zooming there: Understanding nonverbal interaction online

SO YEON PARK, Stanford University, USA

MARK E. WHITING, University of Pennsylvania, USA

Social distancing has shifted us towards virtual platforms, inducing us to engage in video-conferencing at overwhelming rates. Yet, recognizing subtle nonverbal signals in these contexts — being aware of how our actions are presented, and building empathy through video-conferencing — proves challenging as ambiguity increases. We show evidence that unexplained nonverbal actions in online platforms evoke more negative affect than those same actions when their reason is known. Further, through pilots we show that unmanaged nonverbal interaction in video-conferencing can evoke hostile attribution bias. We posit that surfacing rationales behind nonverbal actions will increase empathy and richness in video-conferencing and other online interactions crucial to the future of work. Widespread increases in remote work has amplified the prevalence of these challenges, motivating further study of nonverbal actions online and their consideration in CSCW system designs.

Additional Key Words and Phrases: nonverbal interaction, empathy, teleconference, video chat

1 INTRODUCTION

Communication is mediated by technology more than ever before, and critically, subtle signals valuable to in-person interactions now go unspoken and often unnoticed. Misunderstandings and interpersonal complications arise as these nonverbal signals are increasingly missed online — and not only because our mediums lack fidelity. We study nonverbal interactions in online social platforms, and see that lessons from this research may be informative to the design of computer-supported cooperative work (CSCW) tools and computer-mediated communication (CMC) practices within our accelerating remote work context.

Since its introduction at the World Fair in New York in 1964 by AT&T, video-conferencing has flourished and been studied widely. Seminal works have shown that preference for video-conferencing is palpable, given its high perceived media richness in the context of managerial communication medium selection [2] and its role in enhancing perceptions of productivity from increased communication [4]. An early survey of video-conference room users found that most respondents "liked having regular visual contact with remote collaborators" and many pointed to lack in transmission quality as a negative aspect [32].

Today we have substantially higher quality and more bandwidth in video chat but, by design, some of the cues that are important social signals are now masked due to convenience features [1]. For example, in a recent research workshop discussion, participants anecdotally reflected how others switching video on or off felt like a signal they couldn't understand, potentially indicating boredom, apathy, self-consciousness, or that the 'hiding' conversant seemed to be eating or talking to another person during the video-conference. To others, turning off audio had a similar impact, even though some commented doing so as part of "Zoom etiquette". People were even frustrated by experiences like navigating who should speak or how to end a call without awkward moments of some people waving and others racing to the "Leave" button, only for someone to say one more important thing as half the people drop off the call. It suffices to say, the interaction on video

Authors' addresses: So Yeon Park, syjpark@stanford.edu, Stanford University, Stanford, California, USA; Mark E. Whiting, markew@seas.upenn.edu, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

platforms today is ripe for more intuitive context — we could have more signals about why people do things and what the group intention is at particular points.

Trying to mimic physical co-location through virtual means, and compensating for the lack of proximate interactions has led to a phenomenon popularly called "Zoom fatigue" [11, 15, 24]. "Distance still matters" [21] today, as it did two decades ago, as video-conferencing tools, despite their technological advancements, still cannot help interactants deliver or receive contextual cues and achieve empathy as effectively as they do in-person.

Our findings on nonverbal actions present one promising direction around which video CMC can be improved. We studied nonverbal actions on social platforms with Amazon Mechanical Turk participants. We found that rationales matter in how people perceive these actions and that knowing why someone has performed an action can help recipients feel better about it. These perceptions of nonverbal actions suggest incorporating rationale could help improve affect when using social platforms, including video-conferencing.

In this work, we examine how uncertainty around online nonverbal actions mediates affect and how video chat systems today exacerbate that uncertainty. We show how social platforms need greater clarity of user intentions, and that video-conferencing can also benefit from this insight. We then discuss opportunities to provide improved context for video chat users in a hope to resolve these challenges to improve the future of work.

2   NONVERBAL ACTIONS IN ONLINE PLATFORMS

In collaborative settings, misattributions and misunderstandings often undermine team success and can even cause teams to break apart [35]. This particularly affects creative teams, which rely heavily on these online platforms to exchange ideas, make decisions, and co-create. Inadequate scaffolding of nonverbal actions leads team members to continually question their remote colleagues' behavior, often attributing platform inadequacies to each other rather than to the platform: Why do they not actively contribute to the shared collaborative product even when they seem to be present? Why are they only deleting my content?

The issues of nonverbal actions pervade other online experiences too. One example of the breakdown of such nonverbal online actions occurs when Facebook Likes are over-interpreted as indicating a nuanced social preference. As with team misunderstandings, we get biased towards assuming negative intent in others' neutral actions [17], and attribute someone's negative behavior to their personality, rather than their context or situation [29]. These biases, studied extensively in psychology, are likely more pronounced in technology-mediated interactions because there are fewer, and less effective nonverbal cues on which to base evaluation of a stimulus [3]. Our understanding of nonverbal actions online, analogous to those offline [12], pales in comparison.

To better understand how people respond to nonverbal actions, we designed a multi-phase study focused on action rationales and their effects on recipients. We started by building an understanding of all the possible nonverbal actions that can be taken on social media platforms through task analysis [7] and generative pilots. With these, we focused on the three most used platforms by participant reports and three most common nonverbal actions across our data on those platforms, balancing inherent valence of the actions (positive, neutral, and negative) with Positive and Negative Affect Schedule (PANAS) [34] baseline measures. We then conducted dyadic brainstorming [23] with participants cognizant of the functionalities of all three platforms (and well-versed in at least two) to aggregate a library of plausible rationales for nonverbal actions.

The final instrument studies commenting, liking, and unfollowing, actions present on the three most used social platforms — Facebook, Twitter, and YouTube. Posing nonverbal actions with real user-based rationales enabled us to probe the participants' affect associated with specific reasons. With these materials we assembled a survey instrument based on validated measures

(e.g., I-PANAS-SF [33]) that would ask affect questions given the particular nonverbal action. The question was structured as follows:

> *"Thinking about the receiving person and how they would feel when someone is [performing one of the actions] on [one of the platforms], which the receiving person believes is [because of one of the reasons from dyadic brainstorming sessions],[1] to what extent do you think they would feel:"*
> I-PANAS-SF words were shown with the 5-point Likert scale presented as 'Very slightly or not at all' (1) to 'Extremely' (5).

We used a within-subjects repeated measures design across conditions where reasons were provided — "known" — and one where they were not — "unknown". We instrumented this survey counterbalanced to avoid ordering effects. Through piloting we found the optimal number of nonverbal actions (four action-platform pairs) and rationales (three known reasons) for participants to answer to account for participant fatigue.

We administered our pre-registered survey,[2] to Amazon Mechanical Turk (N = 50) workers who used all three platforms (data including irrelevant text responses were excluded). The survey took approximately 18 minutes to complete, and the respondents were compensated with $4.50 for their participation, achieving a suggested pay rate of $15 per hour [36]. We found a mild trend that knowing the reason for a nonverbal action led recipients to report more positive (and less negative) affect, regardless of the action (Fig. 1). We also asked participants "Do you think knowing the reason for an online action will make you more comfortable with it?", to which 76% of participants responded "Yes, it will make me more comfortable with the online action".

This finding suggests that people's in-situ perceptions of knowing reasons of online nonverbal actions are less impactful than in an imagined situation. This suggests that platforms should embed rationale in mediating user interactions with nonverbal actions in order to help users feel better about receiving nonverbal actions.

3 NONVERBAL ACTIONS IN VIDEO-CONFERENCING

Nonverbal online actions — such as indication of online presence, demarcation of others' contributions, and expressions of support — affect interactants' experiences. Influence of such nonverbal actions — mostly "facial expression, gaze, posture, gesture, and proxemics" [19] — have been studied in the context of video-conferencing. Gestures were found to be used often — to "forecast" their responses, enhance their verbal expression, and convey information non-verbally [10]. Posture and facial expressions were also displayed in video-conferencing to indicate nuanced sentiments such as skepticism, amusement, and conviction [10]. Framing that provided or eliminated some of these gestural and postural cues was also shown to be important, aligning with previous work that facial expressions and bodily gestures are critical to in-person communication [13]; significantly less empathy was found for head-only framing compared to both upper-body framing and face-to-face [19]. Gaze is another powerful nonverbal action that, if channeled, can direct attention faster [22] and can improve trust in group-to-group meetings when spatial faithfulness is captured [18]. However, the efficacy of interactions building on in-person interaction traditions do not tell the entire story.

Video-conferencing can provide a false affordance of visibility [5] — it tricks us into thinking we know more than we actually do about what the other person means. This can lead users toward an illusion of transparency [6] and overestimation of their perceptions, which could be misled. For instance, interviewees rate their interviewers as less attractive, less personable, less trustworthy,

---

[1] In asking participants' affect for actions with unknown reason, we presented the same question without this interpolated clause "which the receiving person believes is [for one of the reasons from dyadic brainstorming sessions]".
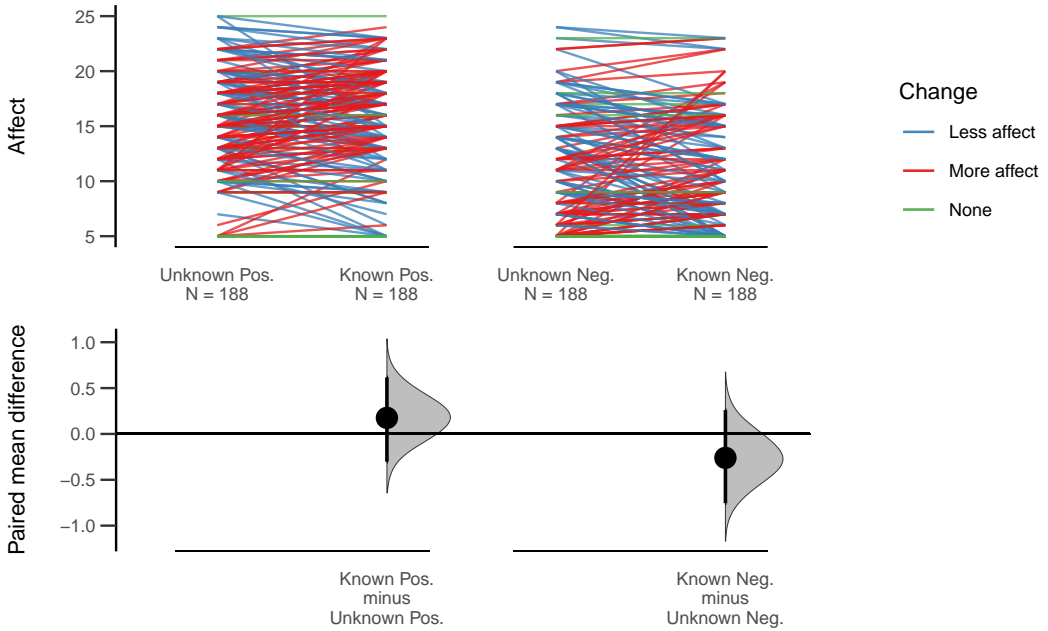[2] https://aspredicted.org/blind2.php

Fig. 1. Cummings plots [8] providing within-subjects comparison of aggregated PANAS affect (broken out by positive and negative affect terms), showing the distribution of means from 5000 bootstrap resamples and 95% confidence intervals (colored by the change in affect of knowing the reason 'known' relative to not knowing the reason 'unknown'). In both cases, knowing reasons for actions improves reported affect — more positive, less negative.

and less competent [27]. Watching the screen intensely could give others the impression that one is concentrating on the meeting, when one may in fact be self-consciously focused on one's own appearance [30] or "[struggling] to find information that just isn't available on screen" [24]. In aggregate, this situation suggests that video chat remains in a perceptual uncanny valley [14] — it works well enough to trick us into thinking it works better than it does, which leads to emotional fatigue.

There are disparities in how people feel about instances of nonverbal actions in video-conferencing. During an academic workshop that we conducted online after the start of the COVID-19 outbreak, attendees unpacked various nonverbal actions taken in Zoom calls and the interpretations behind them. In describing their perceptions of others turning off their camera for Zoom calls, members often attributed these decisions to negative motivations, from not paying attention to the meeting to hiding the fact that they were doing something else. However, in describing why they turned off their own cameras, they gave reasons such as wanting to eat while on the call, briefly attending to other in-person matters (e.g., answering the door or a phone call), and for some privacy (e.g., changing attire, using the restroom). One participant, upon articulating reasons for these nonverbal actions, noticed the possibility of having "double-standards" in considering such nonverbal actions.

We conducted a pilot survey on Amazon Mechanical Turk ($N = 32$) after observing evidence of hostile attribution bias [17] in our workshop between actors and recipients of the nonverbal action of turning the camera off, in order to better understand the divergence of opinions.

We asked participants whether their use of video-conferencing increased with the COVID-19 pandemic, what platform they used most often recently, and how often they engaged in video-conferencing recently. We framed questions as Bayesian truth serum [25] on how they thought people would agree with the following statement: "I have good reasons for turning my video off when video-conferencing", and how positively they thought people would feel about someone turning their video off. Finally, we asked participants to share any final thoughts regarding turning video off in video-conferencing.

90% of these participants answered that their use of video-conferencing increased, 52% used Zoom most often recently, and 62% video-conferenced at least once per week. While the average degree to which participants thought people would agree with the statement *"I have good reasons for turning my video off when video-conferencing"* was above neutral, the levels of agreement were mixed (Fig. 2a). About half of the participants answered "neither positive nor negative" regarding how they thought people would feel about others turning off their video (Fig. 2b).

We found a mix of interpretations of others turning video off, ranging from positive and being more understanding to negative, when asked to share their thoughts about turning the camera off when video-conferencing in general.

> **Positive:**
> (P8) *"I think it's perfectly fine, it's kind of personal actually to have your video camera on, and it should be fine if you don't want it to be on"*
> (P26) *"Some people just don't feel comfortable showing themselves on camera"*
> **Negative:**
> (P21) *"I feel like it's rude because not everyone can do that or there would be no point to video conferencing"*
> (P31) *"it just makes them seem suspicious and less trustworthy when they do that"*

The diverging opinions on turning off video alone shows that it might never be safe for any one person — though possibly safer for those with children in this pandemic as they were often given the benefit of the doubt — to turn off their video, as other participants in the video-conference may perceive their decision negatively.

Moreover, one also stated *"I never really thought it was an option, or considered it as one"* (P12), evidencing the anti-affordance [20] of the video functionality and that the existence of the feature was not enough to make a user feel its affordance, given the social norms and expectations set in place.[3] Such divergent thoughts can form the basis for misunderstandings, and while some participant responses are encouraging, a constant lack of face-to-face could still lead to trust breaking down [28].

These are preliminary results of only one of the nonverbal actions we have mentioned. We foresee similar mixed opinions in other nonverbal interactions, such as how one exits a meeting. Today's video-conferencing interactants experience some discomfort with this (despite prior work on more graceful ways of approaching and leaving [31]) — leaving without saying goodbye during an ongoing meeting can seem impersonal, looking for the "Leave" button can seem awkward, and waving to account for the seemingly-"sterile" goodbye can seem a bit much.

4 CURRENT IMPLICATIONS

Our results shed light on the impact of not knowing why someone performs a nonverbal action. This uncertainty is felt by the recipient in a more negative way than if they had known the reason

---

[3]In explaining anti-affordance, Norman writes that "blockage of passage can be considered an anti-affordance — the prevention of interaction" and notes "pipes obviously blocked cars and trucks from driving on that road" as a good example of such [20].
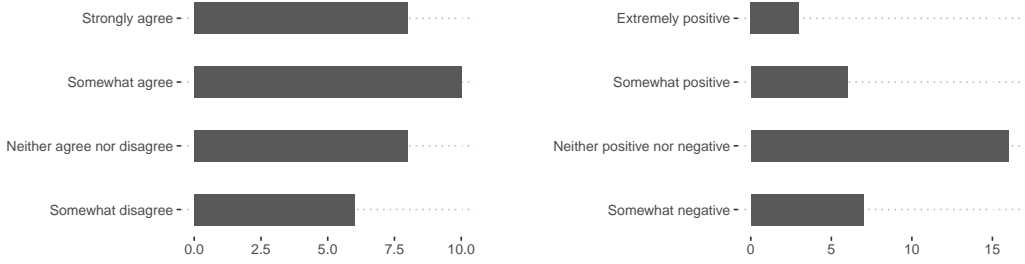
(a) Indication of how much *"people would agree with the following statement: I have good reasons for turning my video off when video-conferencing"* were spread across four of five answer choices.

(b) Answers to *"How positive do you think people feel about others turning their video off when video-conferencing?"* were dominantly neither positive nor negative.

Fig. 2. Pilot study results on video-conferencing, asking questions framed as Bayesian truth serum to increase response accuracy [25], resulted in mixed views on people's feelings towards turning video off ($N = 32$).

behind the action. This suggests hostile attribution bias in online platforms, and necessitates the need to understand how to better design online tools to ameliorate the perceived affect due to this uncertainty.

The importance of knowing the rationale behind a certain action does not end at social media platforms, and also extends to the context of video-conferencing as well — as evidenced from our workshop and our pilot survey. Elucidating the rationale behind why someone may be turning their video off can lead to less negative affect perceived by the observers as well as less hostile attribution bias. Enabling interactants to do so through the design of the platform will also help to ensure the comfort and psychological safety in doing so.

The uncertainty, and the subsequent negative affect, may currently be influencing how users interact with each other, whether on social platforms or in video-conferencing. As we continue to interact with each other via virtual means, users' feelings about others can accumulate for the worse through such uncertain encounters. Such negative (or less positive) emotion accumulation [26] can result in stress for the individual and strain the interactions, thereby inducing a negatively reinforced cycle. Therefore, to enable more sustainable remote interactions, designing for ways to incorporate rationale into our online nonverbal actions will be empowering for users.

The design of nonverbal actions also requires ethical considerations. Today's tools have undermined the communication landscape for some users — they can no longer count on nonverbal signals they expect to see in other settings — and this, we think, is an important reason for more attention to be paid to the careful design of nonverbal actions in tools for work. However, other users may be positively enabled by the current situation — people uncomfortable speaking up in face-to-face interaction sometimes thrive on mediated settings — suggesting that a solution to this situation is not as simple as instituting a specific policy for all users. As a striking example of this issue, some software used for online lectures allows instructors to force all students to enable their video [16], which frustrates students and exacerbates privacy concerns and power imbalances in the interaction. Because of the complex nature of this issue, we identify the design of nonverbal interaction as an area requiring thoughtful consideration and substantial further research — it has the potential to move us beyond reproducing an in-person status quo, but can only achieve that goal with the utmost care.

4.1 Research opportunities

Our work shows that nonverbal actions ought to be better mediated, as they are currently indicating reduced positive affect when interactants do not know the reasons for the actions. Uncovering nonverbal actions that lead to more negative affects as well as studying how to embed rationale into our nonverbal actions are invaluable directions for further research. Understanding how to parse nonverbal actions in more collaborative CSCW and CMC settings also pose exciting areas of study. Here, we have explored one ubiquitous context in which nonverbal actions abound — video-conferencing — and we need to better understand how platform mediation, or lack thereof, in our nonverbal actions affects our interactions. Further, considering various groups of extreme users, such as designers whose practices have moved online despite the inadequacy of current tools for their work processes (e.g., needfinding, empathy-building), will serve to build more tangible insights for video-conferencing tools and other tools advancing remote work. Our findings from social platforms motivate the urgency with which these nonverbal actions ought to be further investigated, as remote work continues to expand and are relied upon.

5 CONCLUSION

Online platforms for work are no longer our future, but our present. As misunderstandings and skewed signals proliferate in CMC, understanding and properly scaffolding nonverbal actions with users' rationale are vital. We have found through a survey-study that providing a rationale can help the recipients to feel better about a nonverbal action taken as uncertainty is reduced. In this paper, we explored one salient nonverbal action in the video-conferencing context, and show how users have divergent views on the action. These divergent views, if left unarticulated and unresolved, can lead to misunderstandings and accumulation of negative emotion, often unbeknownst to the interactants. While online tools, including video-conferencing, have improved greatly, major challenges remain and more can be done to help to achieve "beyond being there" [9], and our work presents clear and compelling opportunities for more positive future of work.

ACKNOWLEDGMENTS

REFERENCES

[1] Ajay Chawla. 2020. Coronavirus (COVID-19)–'Zoom'Application Boon or Bane. *Available at SSRN 3606716* (2020).
[2] Richard L Daft and Robert H Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science* 32, 5 (1986), 554–571.
[3] Alan R Dennis and Susan T Kinney. 1998. Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information systems research* 9, 3 (1998), 256–274.
[4] Stephen Gale. 1990. Human aspects of interactive multimedia communication. *Interacting with computers* 2, 2 (1990), 175–189.
[5] William W Gaver. 1991. Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 79–84.
[6] Thomas Gilovich, Kenneth Savitsky, and Victoria Husted Medvec. 1998. The illusion of transparency: biased assessments of others' ability to read one's emotional states. *Journal of personality and social psychology* 75, 2 (1998), 332.
[7] JoAnn T Hackos and Janice Redish. 1998. *User and task analysis for interface design.* Vol. 1. Wiley New York.
[8] Joses Ho, Tayfun Tumkaya, Sameer Aryal, Hyungwon Choi, and Adam Claridge-Chang. 2019. Moving beyond P values: data analysis with estimation graphics. *Nature methods* 16, 7 (2019), 565–566.

[9] Jim Hollan and Scott Stornetta. 1992. Beyond Being There. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92)*. ACM, New York, NY, USA, 119–125.

[10] Ellen A Isaacs and John C Tang. 1994. What video can and cannot do for collaboration: a case study. *Multimedia systems* 2, 2 (1994), 63–73.

[11] Manyu Jiang. 2020. The reason Zoom calls drain your energy. *Remote Control, BBC. Online: https://www. bbc. com/worklife/article/20200421-why-zoom-video-chats-are-so-exhausting* (2020).

[12] Robert E Kraut, Susan R Fussell, Susan E Brennan, and Jane Siegel. 2002. Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. *Distributed work* (2002), 137–162.

[13] Albert Mehrabian et al. 1971. *Silent messages*. Vol. 8. Wadsworth Belmont, CA.

[14] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19, 2 (2012), 98–100.

[15] Betsy Morris. 27 May, 2020. Why Does Zoom Exhaust You? Science Has an Answer. https://www.wsj.com/articles/why-does-zoom-exhaust-you-science-has-an-answer-11590600269?mod=hp_lead_pos13 Accessed 1 June, 2020.

[16] Michelle Mouton. 2020. Using Zoom to Teach Synchronously Online. *Technology Tools for Teaching in Higher Education, The Practical Handbook Series* (2020).

[17] William Nasby, Brian Hayden, and Bella M DePaulo. 1980. Attributional bias among aggressive boys to interpret unambiguous social stimuli as displays of hostility. *Journal of abnormal psychology* 89, 3 (1980), 459.

[18] David T Nguyen and John Canny. 2007. Multiview: improving trust in group video conferencing through spatial faithfulness. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1465–1474.

[19] David T Nguyen and John Canny. 2009. More than face-to-face: empathy effects of video framing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 423–432.

[20] Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic books.

[21] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human–computer interaction* 15, 2-3 (2000), 139–178.

[22] Mai Otsuki, Keita Maruyama, Hideaki Kuzuoka, and Yusuke Suzuki. 2018. Effects of enhanced gaze presentation on gaze leading in remote collaborative physical tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.

[23] Paul B Paulus and Mary T Dzindolet. 1993. Social influence processes in group brainstorming. *Journal of Personality and Social Psychology* 64, 4 (1993), 575.

[24] Don Pittis. 29 June, 2020. Video chats short circuit a brain function essential for trust — and that's bad for business. https://www.cbc.ca/news/business/zoom-trust-business-pandemic-1.5628638?cmp=rss Accessed 29 June, 2020.

[25] Dražen Prelec. 2004. A Bayesian truth serum for subjective data. *science* 306, 5695 (2004), 462–466.

[26] Maxime Résibois, Philippe Verduyn, Pauline Delaveau, Jean-Yves Rotgé, Peter Kuppens, Iven Van Mechelen, and Philippe Fossati. 2017. The neural basis of emotions varies over time: different regions go with onset-and offset-bound processes underlying emotion intensity. *Social cognitive and affective neuroscience* 12, 8 (2017), 1261–1271.

[27] Fernando Robles, Greg J Sears, Haiyan Zhang, Willi H Wiesner, Rick D Hackett, and Yufei Yuan. 2013. A comparative assessment of videoconference and face-to-face employment interviews. *Management Decision* (2013).

[28] Elena Rocco. 1998. Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 496–502.

[29] Lee Ross. 1977. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology*. Vol. 10. Elsevier, 173–220.

[30] Heather Schwedel. 2 December, 2019. Staring at the Gargoyle on My Screen. https://slate.com/human-interest/2019/12/video-conferencing-is-the-worst.html Accessed 24 June, 2020.

[31] John C Tang. 2007. Approaching and leave-taking: Negotiating contact in computer-mediated communication. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 1 (2007), 5–es.

[32] John C Tang and Ellen Isaacs. 1992. Why do users like video? *Computer Supported Cooperative Work (CSCW)* 1, 3 (1992), 163–196.

[33] Edmund R Thompson. 2007. Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of cross-cultural psychology* 38, 2 (2007), 227–242.

[34] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

[35] Mark E Whiting, Allie Blaising, Chloe Barreau, Laura Fiuza, Nik Marda, Melissa Valentine, and Michael S Bernstein. 2019. Did It Have To End This Way? Understanding the Consistency of Team Fracture. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–23.

[36] Mark E Whiting, Grant Hugh, and Michael S Bernstein. 2019. Fair Work: Crowd Work Minimum Wage with One Line of Code. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 7. 197–206.