A Little Goes a Long Way: How Gesture Visibility in Video Lectures Impacts Attention and Learning

Elena C. Cuffari (elena.cuffari@fandm.edu)
Department of Psychology, Franklin & Marshall College, P.O. Box 3003
Lancaster, PA, 17604 USA

Lauren H. Howard (lauren.howard@fandm.edu)
Department of Psychology, Franklin & Marshall College, P.O. Box 3003
Lancaster, PA, 17604 USA

Ayako Belka (abelka@fandm.edu)
Department of Psychology, Franklin & Marshall College, P.O. Box 3003
Lancaster, PA, 17604 USA

Alonna Reilly (areilly@fandm.edu)
Department of Psychology, Franklin & Marshall College, P.O. Box 3003
Lancaster, PA, 17604 USA

Kyle M. Samson (ksamson@fandm.edu)
Department of Psychology, Franklin & Marshall College, P.O. Box 3003
Lancaster, PA, 17604 USA

Jingyi Sun (jsun3@fandm.edu)
Department of Psychology, Franklin & Marshall College, P.O. Box 3003
Lancaster, PA, 17604 USA

Wei Huang (whuangw@umich.edu)
Department of Sociology, University of Michigan Ann Arbor, 500 S. State St.
Ann Arbor, MI, 48109 USA

Zeyi Zang (zeyi2@illinois.edu)
Department of Psychology, University of Illinois Urbana Champaign, 603 E. Daniel St.
Champaign, IL, 61820 USA

Abstract

In classroom interactions that take place over video conferencing platforms, teachers and students continue to gesture, but their bodies are neither physically copresent nor fully visible to each other. Do instructor gestures help learning in this context, as has been found for in-person learning and for video-based learning in lab experiments? We showed professors lecturing spontaneously with unscripted co-speech gestures. In some conditions, we cropped the video so only the top half of the professor's gesture space is available, or removed the video altogether. Results from our between-subjects experiment show that participants paid significantly more visual attention to the partial gesture condition than to stimuli where the gesturing was fully visible, and they scored significantly higher on an immediate comprehension test if they had seen lectures in the partially visible condition. This work raises further questions of how gestures help learning.

Keywords: online learning; gesture; partial gesture visibility; comprehension; attention; social cognition

Introduction

Prior research shows that the gestures a teacher uses during instruction can positively influence student engagement and learning outcomes. The advent of the Covid-19 pandemic moved many pedagogical interactions to online video-based interaction platforms such as Zoom, where gesture space is limited, and where the classroom community is only virtually, and not corporeally, copresent. Some attribute the benefits of gesturing for learning to the role it plays in establishing and maintaining intersubjectivity in a traditional classroom setting (Majlesi, 2015; Nathan & Alibali, 2011; Nathan et al., 2017). Is it possible to set up spaces of shared meaning-making in the same way in online classrooms? Are gestures still effective here? Though some studies have indicated gesture's positive effects on learning using video stimuli, research (Gullberg & Holmqvist, 1999, 2006) shows that whether in person or while watching video, listeners do not attend visually to speakers' gestures directly, further

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

raising questions as to the mechanisms by which gestures positively impact learning. Given the increasing acceptance of video conference platforms as a classroom setting, we are curious to see just how much gesturing must be visible to have an impact on learning. The present experimental eye-tracking study investigates whether varying conditions of gesture visibility has significant effects on attention and subsequently the comprehension and retention of presented classroom material.

Gestures in Learning

Scholarship has strongly demonstrated that gesture production and observation have positive impacts on learning across the lifespan (Dargue et al., 2019; Hostetter, 2011; Novack & Goldin-Meadow, 2015; Novack et al., 2015; Rueckert et al., 2017). Children are more likely to state that they believe someone who gestures would be a better teacher than someone who doesn't (Wakefield et al., 2021), and they actually learn better from an informant who gestures versus one who does not (Cook et al., 2013; Valenzeno et al., 2003).

Paradigms using video stimuli to compare gesture absence and presence in the same lesson have underscored these findings (Aldugom et al., 2021; Rueckert et al., 2017). It is possible that these effects are driven by aspects of attention; Yang et al. (2023) found that instructor use of deictic gestures in video lectures increased attention in school age participants while undergraduate participants reported greater learning.

While studies have used video as a way of controlling stimuli carefully across conditions, the gesture and not the format (video) is often the focus of these projects. These studies are not designed to capture what happens in online learning, and the gesture conditions are contrasted with no-gesture or incorrect gesture conditions. As lab-based video studies comparing the presence and absence of gesture typically use single, scripted, pedagogical gestures preformatted to match a specific concept (e.g. Rueckert et al., 2017), the particulars of how instructors' spontaneous gestures have effects in video settings are underexplored.

The vast shift towards online meetings in the wake of the Covid-19 pandemic makes the influence of visible gesture space on video a pressing domain to explore. Online learning in general may fall short of education's transformative potential (Maiese, 2013) and of face-to-face learning in key respects (Almahasees et al., 2021; Cranfield et al., 2021). The physical environment in which learning takes place has significant pedagogical implications (Liquin et al., 2023). The "Zoom room" diverges from the traditional classroom (Bailensen, 2021; Lee et al., 2021; McArthur, 2022). All members of a Zoom classroom share the same personal space simultaneously, with no ability to map gaze and the possibility of perpetual self-monitoring. This may contribute to nonverbal overload and subsequent Zoom fatigue (Bailenson, 2021). Faculty also report a struggle to connect in this medium. In McArthur's grounded theory study, 351 instructors were asked to describe their use of various nonverbal communication behaviors in a Zoom classroom, including kinesics (gesture). They frequently commented "…about…the dilemma raised through real-time observation of one's own nonverbal behaviors," with two instructors specifically noting a self-conscious attempt to get their gestures to appear on screen (McArthur, 2022, p. 211).

Interestingly, other research shows that the actual amount of gesturing performed within the visible space of a Zoom room varies from McArthur's (2022) instructor self-reports. A qualitative analysis of 1497 gestures found that the majority deemed semantically relevant to the lecture were partially invisible due to the Zoom frame (71.6% semantically relevant gestures were either cut off at the stroke point or moved on and off the screen, as opposed to 25.7% fully visible semantically relevant gestures) (Cuffari, 2022).

Gesture and Attention

In seeking to determine what drives gesture fixation, Gullberg and Holmqvist conducted several gaze-tracking studies and found that listeners rarely deviate from looking at speaker's faces (Gullberg & Holmqvist, 1999, 2006). Comparing in-person to video conditions, Gullberg and Holmqvist (2006) found fewer gaze fixations on gesture in video conditions than in live, but found overall very few gesture fixations (less than 0.5% of the time) compared to face (over 90% of the time). They also suggest that viewers treat life-size video projection similarly to live interactors, compared to the smaller dimensions of video, which render most gestures visible/detectable without looking away from the face (ibid).

This work raises a quandary: How do students benefit from teachers' gestures if they rarely look at them? If students are gleaning gestural information via peripheral vision, will they then miss out on this information when gestures are cut off?

How Do Gestures Help Learning?

A variety of explanations exist for the mechanisms responsible for gesture's positive influence on learning and recall. For example, gesture helps as a channel of additional information (Singer and Goldin-Meadow, 2005); it captures and maintains attention, provides redundancy in content, and grounds speech in a physical environment (Valenzeno et al., 2003). Kelly et al. (2010) give empirical evidence for gesture and speech operating as an integrated system in comprehension, based on faster processing times when speech and gesture are congruent. Wakefield et al. (2018) find that gesture interacts with speech complexly to support a kind of looking that gleans key information from the speech stream. Goldin-Meadow and Alibali (2013) point to social mechanisms including how listeners treat speakers' gestures as constructions of their current knowledge.

Given our readiness-to-interact (Di Paolo & De Jaegher, 2012), properly social interactions can happen over videoconferencing (Vidolov, 2024), but they are likely to be shorter and in near-continuous breakdown or recovery (Di Paolo et al., 2018). Co-speech gestures in everyday interactions often serve as key examples of the coordination dynamics that can constitute social cognition (Cuffari, 2012;

Cuffari, 2024; De Jaegher & Di Paolo, 2007). It may be that our histories and habits of in-person conversations aid us in making sense of gestures over a video connection, perhaps even of completing them in our own bodies (Merleau-Ponty, 2013), especially when cut off visually. Measuring emergent social dynamics of video mediated interactions is challenging. Osler and Zahavi (2022) insist that "digital encounters constitute their own forms of sociality" and call for their own investigations (see Van Dijk, 2022). In this study, we chose to explore visual attention to spontaneous gestures in a controlled experiment as a starting point.

The literature does not point to a simple prediction about the effects of partial gesture visibility on learning. On the one hand, the presence of gesture in lecture is shown to improve comprehension, particularly of abstract, college-level concepts (Aldugom et al., 2021; Rueckert et al, 2017;). On the other, it is unclear if listeners need to see all of a gesture to reap the benefits (Gullberg & Holmqvist, 2006). Assuming gestures made at the periphery of the Zoom frame would attract attention (Gullberg & Holmqvist, 1999), we hypothesized that partially visible gestures would receive more gaze fixation than fully visible gestures. We also hypothesized that looks to cut-off gestures would prove distracting, lowering comprehension scores; scores should be higher in the full gesture condition where participants can reap the benefits of visible representational gesturing. As gestures have been shown to improve consolidation of learned concepts over time (Cook et al., 2013), we expected the full gesture condition to show higher comprehension scores after a delay vs. during immediate recall.

Methods

Participants

Participants were recruited via introductory psychology courses and the general student population at a small liberal arts college. Participant sample sizes were determined via a combination of prior literature review (e.g., Aldugom et al., 2021; Novack et al., 2015; Sweller et al., 2023) and G*Power analysis (one-way ANOVA, three groups, 0.4 effect size), suggesting a minimum sample size of 102 individuals. The final sample included 94 participants (59 female, 29 male, 4 non-binary/third gender, and 1 other). Participants self-reported as White (59.6%), Asian (20.2%), Black/African American (6.4%), Hispanic/Latino (5.3%), more than one race/ethnicity reported (2.2%), and African, North African, Middle Eastern, Eurasian, and Other each (1.1%). 89.5% of participants listed English as their first language. Participants signed up via SONA systems software (Fidler, 2002) or were recruited via convenience sampling, and elected either to receive course research credit or enter a gift card lottery for participating. Participants were randomly assigned to one of three possible conditions: full gesture (FG), partial gesture (PG), or no gesture (NG).

Materials

Video Stimuli Five stimulus videos depicted professors lecturing in their area of expertise while standing in front of a white board (see Figure 1). All instructors are the same distance from the camera. Professors varied in relation to gender (3 female, 2 male) and discipline (Biology, Economics, Philosophy, Chemistry) and were instructed to present a concept that they teach often, without using props, slides, or a white board. Without any mention of gesture, we simply asked the professors to talk to a small student audience present at the time of the recording, and to avoid covering their mouth while speaking. All professors gestured nearly continuously throughout their presentations. Coding based on McNeill (1992) classified gestures according to type; we collapsed iconic and metaphoric gestures into a representational category we then used to devise comprehension questions (see below).

After recording, the videos were edited such that each was approximately the same length (*M*=179.2 seconds). Additionally, a version of each video was created to correspond to each of the three gesture conditions (see Figure 1). In the full gesture videos (FG), the professor's whole body above the waist could be seen. In the partial gesture videos (PG), the video was cropped roughly at chest level, above common gesture space (McNeill, 1992). To crop the videos we simply placed black boxes over parts of the video, so that scale and face location was unchanged across conditions. In the no gesture videos (NG), all visual aspects were removed leaving a black box with the original audio track.

Each participant was assigned to one of three conditions and watched two stimulus videos (one female and one male professor). The full counterbalancing of these videos resulted in 12 possible video orders for each of the three conditions (thus, 36 possible order/condition combinations altogether).

Tobii x3-120 Eye Tracking System Video stimuli were presented on a 25 inch LCD monitor equipped with a Tobii x3-120 corneal reflection eye-tracking system (accuracy 0.4°, sampling rate 120 Hz, Tobii Technology, Stockholm, Sweden). Stimuli were presented and data was collected using Tobii Studio software (version 3.4.8, Tobii Technology, Stockholm, Sweden).

Connectedness Questionnaire An 18-item questionnaire was created to assess participant interest and engagement with the professors and content in the stimulus videos. This served as the foil for the study, so that participants would not be primed to attend to faculty gesturing during the stimulus presentation or self-consciously to their own gesturing during the comprehension check. Each statement was followed by a 5-point Likert scale where participants could indicate how much they agreed with that statement (1= "strongly disagree", 5="strongly agree"). Questions were equally split between aspects of participant attitude towards the content (e.g., "I feel confident I can explain this topic to someone else") and their attitude towards the professor's ability and approachability (e.g., "I would feel comfortable talking to this professor one on one"). Participants were asked about their familiarity with the professors in the video and the topic in the video in order to account for individual differences in prior knowledge. The questionnaire was administered via Qualtrics, and each participant filled it out once for each video they saw.

Comprehension Test Three comprehension questions were generated for each video to assess each participant's learning and understanding of the video content. The questions and correct answers were created in consultation with the relevant professors when necessary. Two of the questions corresponded to content where the professor was both lecturing and gesturing representationally in the video (e.g., talking about raising prices while raising their hand to show level increase). The third question was an overarching summary question regarding the lecture's main idea. All questions were open-ended and designed to elicit detailed responses, and ideally accompanying gestures, from participants during recall. The comprehension survey was administered once in person and once after a one-week delay via Qualtrics.

Procedure

Upon arrival at the lab, participants completed informed consent and a brief demographic form. They were then moved to an area in the lab to complete the encoding phase.

Encoding Phase Participants sat on a chair approximately 65 cm/25.5 inches in front of the Tobii eye-tracking system. Each participant completed a standard Tobii calibration procedure (minimum 5-point calibration) and then watched two of the stimulus videos, varying according to assigned order and condition (FG, PG, NG). The condition assignment was consistent within participants, but varied across participants, such that each participant would only see videos with full, partial, or no gesture. A small attention-getter with a sound appeared in the middle of the screen between videos in order to re-orient attention and visually inspect for eye tracker drift post-hoc. The encoding phase lasted approximately eight minutes.

Delay Phase After watching the lecture videos, participants were moved to a table across the room where they were asked to complete the Connectedness Questionnaire, presented via Qualtrics on a laboratory laptop. This phase lasted approximately three minutes.

Immediate Recall After finishing this questionnaire, the experimenter verbally asked the participants the three questions from the comprehension survey to test immediate recall. Participants were given as much time as they needed to answer the questions and were reminded that it was acceptable for them to not know the correct response. A camera was positioned in front of the table facing the participant to record their verbal answers and gestures. Upon completion of the test, participants were asked for their contact information and were reminded that they would be sent a repeat comprehension test the following week.

Delayed Recall Exactly one week (7 days) after completing the immediate recall, participants were emailed a Qualtrics form with the same comprehension questions from the immediate recall phase. Participants were also asked if they believed their responses got better the second time they answered these questions, and if they had looked up any answers for these questions at any point during the intervening week. Participants were given 72 hours to complete these questions via Qualtrics.

Coding and Analyses

Eye Tracking Data Attentional data was aggregated and exported for further analyses using Tobii Studio (Tobii Technology, Stockholm, Sweden. Areas of Interest (AOIs) were created for the whole screen, the professor's face, the professor's below-face gesture space, and the professor's face-level gesture space on either side (Figure 1). For the PG condition, an AOI was created called "offscreen" which involved the space where a below-face gesture would have been presented if the video had not been cropped. The total duration of looking to these AOIs (in seconds) was exported for each video and averaged across participants.

Because of the variation in professor height and build, individual AOIs were crafted for each video. Percentiles were then created for each AOI ((duration of attention to the AOI / duration of attention to the whole screen) * 100) in order to account for these differences.

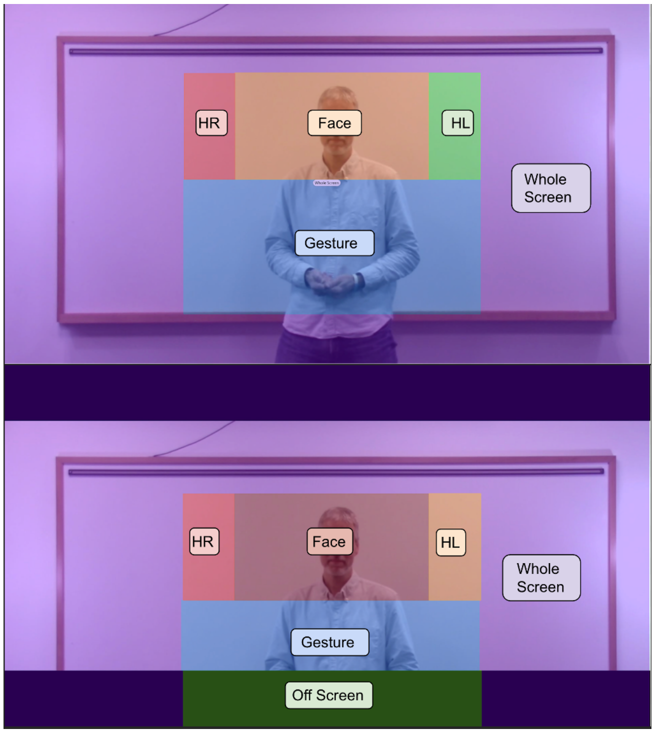

Figure 1: Top image shows full gesture condition (FG), lower image shows partial gesture condition (PG), for the same time point in a stimulus lecture. Top image shows areas of interest for a FG video (HR, Face, HL, Whole Screen, and Gesture); bottom image shows areas of interest for a PG video (HR, Face, Whole Screen, Gesture, and Off Screen).

Comprehension Survey Participant responses were coded on a 0-200 scale for each of the two time points (immediate and delayed). Each of the two content questions could receive a possible score of 0, 50, or 100. The summary question was utilized as a quality check and was not scored. Partial credit (50) was given for responses that answered only parts of the question or expressed general ideas without demonstrating adequate understanding, specifically of the concepts associated with key gestures .

Results

Preliminary analyses found that professor gender, participant gender, and the native language of the participant had no significant effect on any measures (all *p*s >.05). Subsequent analyses were collapsed across these factors.

Attention to Video Stimuli

A one-way ANOVA found a significant difference in overall looking time (in seconds) to the screen based on condition (*F*(2,87) = 11.98, *p*<.001) (Figure 2). Independent sample t-tests revealed that PG participants attended to the whole screen significantly longer (*M*= 254.62 seconds, *SD*= 65.15) than those in FG (*M*= 206.60 seconds, *SD*= 68.46), *t*(57)=-2.76, *p*=.008. Though more information was present on the screen in the FG condition, the PG condition garnered more visual interest. NG participants (*M*=161.49 seconds, *SD*= 86.76) looked at the screen significantly less than FG participants *t*(58)=2.23, *p*=.03, and significantly less than PG participants *t*(59)=4.73, *p*<.001. This was expected, given that no visual information was on the screen in the NG condition. The NG condition was removed from subsequent AOI analyses because participant gaze in this condition was not a reliable indicator of attention to gesture.

Condition also influenced to which areas of the screen participants attended. A series of independent t-tests revealed that participants in the PG condition looked to the *face* AOI (*M*=92.91%, *SD*=0.07) for a significantly greater percent of time compared to participants in the FG condition (*M*=82.92%, *SD*=0.13), *t*(56)=-3.68, *p*<.001. FG participants looked to the *gesture* AOI (*M*=13.95%, *SD*=0.10) for a significantly greater percent of time compared to PG participants (*M*=4.95%, *SD*=0.05), *t*(57)=4.43, *p*<.001. No significant differences were found in percent of looking to the head-left or head-right AOIs across conditions (all *p*s >.05).

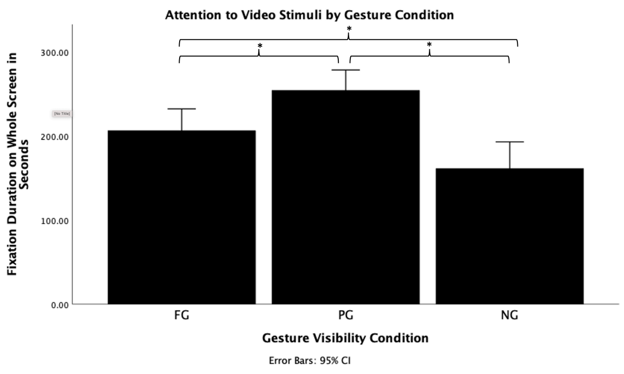

Figure 2: Average attention to video stimuli by gesture visibility condition. Attention was measured as average gaze fixation duration across visibility conditions.

Comprehension Test Performance

One-way ANOVAs were run to explore the influence of gesture visibility on immediate and delayed comprehension of the lectured material (Figure 3). We found a significant difference in immediate comprehension scores across conditions *F*(2)=5.82, *p*=.004. No significant difference was found between comprehension scores at the delayed time point *F*(2)=2.92, *p*> .05. T-tests revealed that PG participants (*M*=239.66 points, *SD*=93.90) scored significantly higher on the immediate comprehension questions than FG participants (*M*=158.06 points, *SD*=114.82), *t*(58)=-3.00, *p*=.004.

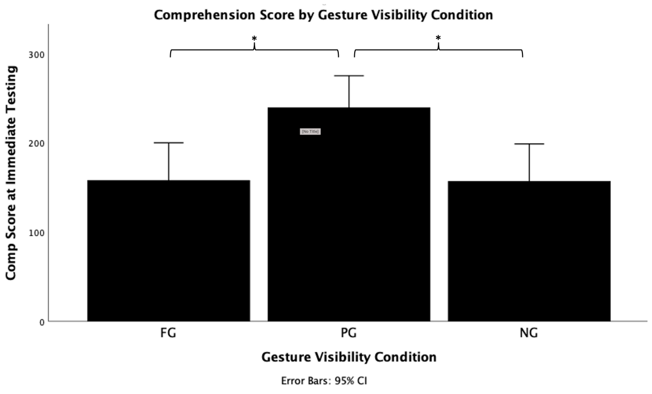

Figure 3: Average immediate recall comprehension score by gesture visibility condition.

Difference scores were created to examine the amount of change in comprehension score from the immediate to delayed test period (difference score = immediate score - delayed score). Planned t-tests found that FG participants had an average comprehension score change of 8.33 from the immediate to the delay phase (*M*=-8.33 points, *SD*=63.70), significantly different from the PG participants (*M*=46.15 points, *SD*=109.47), *t*(48)=-2.127, *p*=.039. Performance change for PG participants was an average of 46 points lower, and performance change was worse for NG participants by an average of 37 points (*M*=37.04 points, *SD*=75.44).

Of note, the patterns noted above remained significant even after excluding participants who said they had prior knowledge of the lectured subject.

Connection Between Comprehension and Attention

A series of Pearson's correlations were run to examine relationships between attention to the overall screen, the AOIs, and comprehension scores. When NG cases were excluded, there was a positive correlation between the immediate comprehension test score and the average percent of time looking to face, *r*(54)=.38, *p*=.002. There was a negative correlation between the immediate comprehension test score and the average percent of time looking to gesture, *r*(55)=-.38, *p*=.002. There was a positive correlation between the comprehension difference score and the average percent of time looking to face, *r*(45)=.29, *p*=.026. There was a negative correlation between the comprehension difference score and the average percent of time looking to gesture, *r*(46)=-.29, *p*=.022.

We then examined these correlations by condition. In the FG condition there was a significant positive correlation between looking to the face AOI and immediate comprehension score, *r*(26)=.34, *p*=.04. In the PG condition there was a significant positive correlation between looking to the face AOI and the comprehension difference score, *r*(23)=.39, *p*=.027. This indicates that the more participants in this condition looked to the face AOI, the greater *negative* change they had in scores between immediate and delay time points. When examined by condition, no significant correlations were found between comprehension scores and attention to the gesture AOI, nor between comprehension scores and attention to the whole screen or other AOIs.

Discussion

To explore the role of gesturing in learning in online classrooms, we altered the visibility of a professor's primary gesture space (McNeill, 1992) and examined the effect on content comprehension. In our stimuli professors gestured freely in the course of explaining a concept they often teach. This allowed us to test comprehension precisely at the moments when key representational gestures were or were not fully visible.

Per studies demonstrating the beneficial effects of representational and deictic gestures in conceptual learning and in narrative recall (e.g., Dargue et al., 2017; Rueckert et al., 2019), we hypothesized that comprehension scores would be highest in the full gesture condition. They were, in fact, lowest in this condition. Participants paid significantly more visual attention to the *partial* gesture condition than to stimuli where the gesturing was fully visible, and they scored significantly higher on the comprehension test if they had seen lectures in the PG condition as opposed to full or no gestures.

In PG, subjects looked significantly more at the face than those in the FG condition; they did not ‘go looking for’ the rest of the partially absent gestures. Gaze fixation duration overall and on the face were both significantly higher in PG, suggesting that participants paid more focused attention in this condition. This then raises the question of why the PG condition would be better at holding participants’ focused attention.

Although we had suspected that cut off gestures could prove distracting to viewers, Gullberg and Holmqvist (2006) suggest that gesturing while speaking is equivalent to visual background ‘noise,’ and draws more attention when it stops. It could be that viewers registered the partial gestures as typical speech-accompanying movement and, in the absence of the possibility of focusing on peak strokes, holds, or far-flung peripheral gestures clearly and fully, narrowed their attention to the face and to the content. It is also possible that watching the full gesture videos increased the participants’ cognitive load, and that here, in fact, gesturing proved distracting (Moon & Ryu, 2021).

Viewers in the FG condition attended to the gesture AOI 14% of the time on average, quite a bit more than Gullberg and Holmqvist’s finding of about .05% fixation on hand gesturing in both live and video conditions (1999, 2006). Ample attention to gesturing in the FG condition could explain the improvement on performance in the delayed comprehension task, based on Cook et al. (2013). However, the overall poorer comprehension performance in this condition flags the need for further investigation of gestures’ beneficial effects on learning when gesturing is spontaneous rather than pre-designed.

While there were significant differences in both attention and comprehension across conditions, we found only two significant relationships between attention to parts of the screen and later memory scores. These accord with our reported results, insofar as looking to the face is helpful in the immediate comprehension test but yields a negative change in comprehension over time, whereas participants with average greater looking to gesture time improved comprehension over time.

Our unexpected findings prompt the need for further and improved testing in this novel paradigm. With interactive labor reduced and given visual constraints of the Zoom frame, at least in institutional contexts like the undergraduate college (online) classroom, perhaps participants focus on content. This narrowing of the attentional field may then lead to better retention and testing performance, initially. Longer-term learning as measured by essay writing for example require further study. The current study serves as an important first step in exploring the role of partial (cut off) gesture on learning, setting the foundation for future studies that could explore these questions in live interactive contexts.

Acknowledgments

We would like to thank the Hackman Summer Scholar Fund, the Richard C. Puff Fund, and the Committee on Grants grant to the first author at Franklin and Marshall College for funding this project.

References

Aldugom, M., Fenn, K., Day, A., & Cook, S. W. (2021). The Role of Verbal and Visuospatial Working Memory in Supporting Mathematics Learning With and Without Hand Gesture. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*(43).

Almahasees, Z., Mohsen, K., & Amin, M. O. (2021). Faculty's and Students' Perceptions of Online Learning During COVID-19. *Frontiers in Education*, *6*. https://www.frontiersin.org/articles/10.3389/feduc.2021.638470

Bailenson, J. N. (2021). Nonverbal Overload: A Theoretical Argument for the Causes of Zoom Fatigue. *Technology, Mind, and Behavior, 2*(1). https://doi.org/10.1037/tmb0000030

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and Transfer of Learning After Observing Hand Gesture. *Child Development*, *84*(6), 1863–1871. https://doi.org/10.1111/cdev.12097

Cranfield, D. J., Tick, A., Venter, I. M., Blignaut, R. J., & Renaud, K. (2021). Higher education students' perceptions of online learning during COVID-19—A comparative study. *Education Sciences*, *11*(8), 403. https://doi.org/10.3390/educsci11080403

Cuffari, E. (2012). Gestural sense-making: hand gestures as intersubjective linguistic enactments. *Phenomenology and the Cognitive Sciences*, *11*, 599-622. https://doi.org/10.1007/s11097-011-9244-9

Cuffari, E. (2022, December 9). *Partial gestures, partial languaging?: Searching for participatory sense-making in video-mediated classrooms.* [Invited paper]. ISGS Hong Kong: A Day of Sense-Making: Living, Moving, Knowing, Feeling, Gesturing, Hong Kong.

Cuffari, E. (2024). Intersubjectivity and Gesture. In *Cambridge Handbook of Gesture Studies*. A. Cienki (Ed.), (pp. 599-615). Cambridge University Press.

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin*, *145*(8), 765. http://dx.doi.org/10.1037/bul0000202

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*, 485-507. 10.1007/s11097-007-9076-9

Di Paolo, E., & De Jaegher, H. (2012). The interactive brain hypothesis. *Frontiers in Human Neuroscience*, *6*, 163. https://doi.org/10.3389/fnhum.2012.00163

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language*. MIT press.

Gullberg, M., & Holmqvist, K. (1999). Keeping an eye on gestures: Visual perception of gestures in face-to-face communication. *Pragmatics & Cognition*, *7*(1), 35–63. https://doi.org/10.1075/pc.7.1.04gul

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, *14*(1), 53–82. https://doi.org/10.1075/pc.14.1.05gul

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin*, *137*(2), 297–315. https://doi.org/10.1037/a0022128

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*(2), 260–267. https://doi.org/10.1177/0956797609357327

Lee, A. Y., Moskowitz-Sweet, G., Pelavin, E., Rivera, O., & Hancock, J. T. (2021). "Bringing you into the Zoom": The power of authentic engagement in a time of crisis in the U.S. *Journal of Children and Media*, *15*(1), 91–95. https://doi.org/10.1080/17482798.2020.1858437

Liquin, E. G., Luzuriaga, N., & Gureckis, T. M. (2023). Teaching and Learning Through Pedagogical Environment Design. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45). https://escholarship.org/uc/item/9xq3w7rc

Maiese, M. (2013). Embodied social cognition, participatory sense-making, and online learning. *Social Philosophy Today, 29*, 103-119. https://doi.org/10.5840/socphiltoday201329111

Majlesi, A. R. (2015). Matching gestures–Teachers' repetitions of students' gestures in second language learning classrooms. *Journal of Pragmatics*, *76*, 30–45. https://doi.org/10.1016/j.pragma.2014.11.006

McArthur, J. A. (2022). From classroom to Zoom room: Exploring instructor modifications of visual nonverbal behaviors in synchronous online classrooms. *Communication Teacher*, *36*(3), 204–215. https://doi.org/10.1080/17404622.2021.1981959

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

Merleau-Ponty, M., Landes, D., Carman, T., & Lefort, C. (2013). *Phenomenology of perception*. Routledge.

Moon, J., & Ryu, J. (2021). The effects of social and cognitive cues on learning comprehension, eye-gaze pattern, and cognitive load in video instruction. *Journal of Computing in Higher Education*, *33*(1), 39–63. https://doi.org/10.1007/s12528-020-09255-x

Nathan, M. J., & Alibali, M. W. (2011). How gesture use enables intersubjectivity in the classroom. *Integrating Gestures*. https://library.oapen.org/bitstream/handle/20.500.12657/45671/625251.pdf?sequence=1#page=266

Nathan, M. J., Alibali, M. W., & Church, R. B. (2017). Making and breaking common ground. *Why Gesture?: How the Hands Function in Speaking, Thinking and Communicating*, *7*, 285.

Novack, M. A., Goldin-Meadow, S., & Woodward, A. L. (2015). Learning from gesture: How early does it happen? *Cognition*, *142*, 138–147. https://doi.org/10.1016/j.cognition.2015.05.018

Novack, M., & Goldin-Meadow, S. (2015). Learning from gesture: How our hands change our minds. *Educational*

Psychology Review, *27*(3), 405–412. https://doi.org/10.1007/s10648-015-9325-3

Osler, L., & Zahavi, D. (2022). Sociality and Embodiment: Online Communication During and After Covid-19. *Foundations of Science*. https://doi.org/10.1007/s10699-022-09861-1

Rueckert, L., Church, R. B., Avila, A., & Trejo, T. (2017). Gesture enhances learning of a complex statistical concept. *Cognitive Research: Principles and Implications*, *2*(1), 1–6.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological science*, *16*(2), 85-89. https://doi.org/10.1111/j.0956-7976.2005.00786.x

Sweller, N., Choi, A.-J., & Austin, E. (2023). Gesture production at encoding supports narrative recall. *Psychological Research*. https://doi.org/10.1007/s00426-023-01886-w

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*(2), 187–204. https://doi.org/10.1016/S0361-476X(02)00007-3

Van Dijk, J. (2022). Zooming in on embodied social sensemaking: Mapping the design space in the context of videoconferencing. *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction*, 1–10. https://doi.org/10.1145/3490149.3501324

Vidolov, S. P. (2024). Virtual collaboration as co-enacting intercorporeality. *European Journal of Information Systems*, *33*(2), 244-266. https://doi.org/10.1080/0960085X.2022.2152743

Wakefield, E., Novack, M. A., Congdon, E. L., Franconeri, S., & Goldin-Meadow, S. (2018). Gesture helps learners learn, but not merely by guiding their visual attention. *Developmental Science, 21*(6), e12664. https://doi.org/10.1111/desc.12664

Wakefield, E. M., Novack, M. A., Congdon, E. L., & Howard, L. H. (2021). Individual differences in gesture interpretation predict children's propensity to pick a gesturer as a good informant. *Journal of Experimental Child Psychology*, *205*, 105069. https://doi.org/10.1016/j.jecp.2020.105069

Yang, J., Zhu, F., Jiang, Y., & Pi, Z. (2023). Do adults and children learn differently from video lectures with an instructor's deictic gestures?. *Education and Information Technologies*, *28*(7), 8377-8400. https://doi.org/10.1007/s10639-022-11523-5