



# Perception of being observed by a speaker alters gaze behavior

Michael J. Kleiman<sup>1</sup> · Elan Barenholtz<sup>2</sup>

Published online: 18 February 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

Previous research has shown that gaze behavior of a speaker's face during speech encoding is influenced by an array of factors relating to the quality of the speech signal and the encoding task. In these studies, participants were aware they were viewing pre-recorded stimuli of a speaker that is not representative of natural social interactions in which an interlocutor can observe one's gaze direction, potentially affecting fixation behavior due to communicative and social considerations. To assess the potential role of these factors during speech encoding, we compared fixation behavior during a speech-encoding task under two conditions: in the "real-time" condition, we used deception to convince participants that they were interacting with a live person who was able to see and hear them through online remote video communication. In the "pre-recorded" condition, participants were correctly informed they were watching a previously recorded video. We found that participants fixated the interlocutor's face significantly less in the real-time condition than the pre-recorded condition. When participants did look at the face, they fixated the mouth at a higher proportion of the time in the pre-recorded condition versus the real-time condition. These findings suggest that people engage in avoidance of potentially useful speech-directed fixations when they believe their fixations are being observed and demonstrate that social factors play a significant role in fixation behavior during speech encoding.

**Keywords** Eye movements · Fixation behavior · Simulated dyadic interaction · Pre-recorded video

## Introduction

Previous research on fixation behavior during speech encoding has found that the dynamic face of the speaker attracts the observer's attention, with the mouth and eyes serving as the primary – and competing – sites of fixation. The mouth tends to attract more attention when the auditory signal is degraded (Buchan, Paré, & Munhall, 2007; Lansing & McConkie, 2003; Vatikiotis-Bateson, Eigsti, Yano, & Munhall, 1998) or the language is unfamiliar (Barenholtz, Mavica, & Lewkowicz, 2016). However, when social reference, emotional, or deictic cues are relevant the eyes attract more fixations, particularly when determining intentionality (Birmingham, Bischof, & Kingstone, 2008; Buchan, Paré, &

Munhall, 2008; Emery, 2000; Vö, Smith, Mital, & Henderson, 2012).

To date, research on fixation behavior during speech-encoding tasks has only considered cases in which participants know that they are viewing pre-recorded stimuli and therefore are aware that the speaker that they are viewing cannot see where they themselves are fixating, as they would in a real-world interaction. However, knowing that one's gaze is visible to others – as is typical of real-life interactions – is likely to affect fixation behavior. This is because the direction of one's gaze not only serves to encode information from an interlocutor but can also serve social and communicative roles as well. People are highly sensitive to the gaze direction of others; infants as young as 2 days old show a preference for faces in which the eyes are looking directly back at them (Farroni, Csibra, Simion, & Johnson, 2002; Farroni, Menon, & Johnson, 2006) and infants (as well as adults) show a preference for looking at the eyes, relative to other parts of the face (Maurer, 1985). These early tendencies have been proposed to reflect evolved mechanisms for detecting eye-gaze direction (Baron-Cohen, 1995). Humans are unique among primates in having a white sclera, which increases contrast with the pupil, allowing easier discrimination of gaze direction (Kobayashi & Kohshima, 2001). Importantly, while other species show

---

✉ Michael J. Kleiman  
michael@kleiman.me

<sup>1</sup> Department of Psychology, Florida Atlantic University, Boca Raton, FL, USA

<sup>2</sup> Department of Psychology/Center for Complex Systems and Brain Sciences, Florida Atlantic University, 777 Glades Rd., Boca Raton, FL 33433, USA

sensitivity to being fixated (Burghardt & Greene, 1990; Scaife, 1976), humans are likely unique in that they reflexively allocate attention to the perceived gaze direction of others, a phenomenon known as “gaze cueing” (Driver et al., 1999; Friesen & Kingstone, 1998; Langton & Bruce, 1999; although see Tipples, 2002). This mechanism likely plays a role in the developmentally and socially important phenomenon of “shared” or “joint” attention in which multiple people attend to the same object or location (Bruner & Sherwood, 1983; Scaife & Bruner, 1975). Interestingly, this tendency is greatly reduced in children with autism spectrum disorder (Baron-Cohen, 1989; Baron-Cohen, Allen, & Gillberg, 1992).

Because gaze direction conveys so much socially relevant information, one's own gaze behavior is likely to be affected by whether one's eyes are visible to an interlocutor. For example, people may intend to signal that they are paying attention to an interlocutor by fixating their face or eyes during a conversation. Conversely, extended eye contact can also be perceived as aggressive (Nichols & Champness, 1971) and therefore the observability of one's eyes could lead to *reduced* direct fixation of another's face or eyes. Indeed, people engage in avoidant eye movements – the periodic breakage and reformation of eye contact during conversations (Griffin & Bock, 2000; Ho, Foulsham, & Kingstone, 2015; Morency, Christoudias, & Darrell, 2006). This tendency has been found to increase with cognitive load when thinking (Glenberg, Schroeder, & Robertson, 1998) or when encoding new information (Lusk & Mitchel, 2016; Spivey & Geng, 2001) and has both social and cultural dimensions as well; avoidance increases when participants believe that video of them will later be reviewed by someone of higher social rank (Gobel, Kim, & Richardson, 2015) and for those in East Asian cultures compared to Western cultures (Lee, Greene, Tsai, & Chou, 2016; Senju et al., 2013).

In the current study, we compared fixation behavior when people engaged in a speech-encoding task under two conditions; a “real-time” condition in which participants were led to believe they were engaging in a real-time two-way video interaction where they could be seen and heard by the speaker, and a “pre-recorded” condition in which participants were made aware that the video was previously recorded and therefore the speaker could not see their behavior. We hypothesized that participants would alter their fixation behavior between the real-time and pre-recorded conditions with several possible outcomes: face fixation may increase in the real-time condition based on the social expectation of facing one's interlocutor in order to demonstrate attention, however it is also possible that the real-time condition will lead to greater face avoidance, based on social norms as well as the cognitive demands of encoding the lecture. Similarly, with regard to where on the face the participant may fixate, it is possible that participants will fixate the eyes more in the real-time condition because of social demands to make eye-contact with one's

interlocutor. Inversely, in the pre-recorded condition, where the social demands to make eye contact are eliminated, participants may spend more time looking at the mouth in order to encode the lecture, consistent with previous studies showing greater mouth fixations during an encoding task.

## Method

### Participants

One hundred and seventy-three undergraduate students, enrolled in an introductory psychology course at Florida Atlantic University, participated in the study for course credit. All participants spoke at least fluent English and reported having normal or corrected-to-normal vision and hearing. Participants were naive to the purposes of the experiment. The majority of our participants were born and raised in Western countries (93/98 participants), diminishing our ability to examine cultural background as an influence. Informed consent was collected for all participants.

### Stimuli

Four videos were used in the experiment, each consisting of one model reading one of two lectures. Two models were used, one male and one female. The actors read the script off of a teleprompter, with the camera positioned approximately 2 feet in front of the teleprompter in order to minimize visible eye movements as well as to promote the feeling of the actors making eye contact with the camera. The models were recorded from the shoulders up, with a neutral emotion and prosody, and with a neutrally white backdrop.

### Procedure

To investigate the social effects of fixation behavior, we developed a paradigm that encouraged participants to believe that they were participating in a real-time interaction despite all stimuli being pre-recorded. Prior to being shown the first video, participants were informed that they would be participating in an experiment to examine learning and memory while communicating via real-time online conferencing software. They were also told that the person they would be speaking to is located on another campus, and to ignore any delay that may be caused by the internet. Additionally, a point was made by the experimenter to pretend to text the actor, while stating “let me make sure the other person is ready for you,” in order to further ensure the deception was successful.

In the real-time condition, after calibration was performed and the experimenter had left the room, a “dialing” sequence was displayed to the participant followed by a prompt to “Press any key to accept the call,” which launched the real-time stimulus (Fig. 1A). An “interactive” portion was then



**Fig. 1** Experimental sequences with durations shown in timeline (Note: A–C are only shown in the “real-time” condition). (A) *Dialog boxes mimicking a dialing sequence.* (B) *Call is answered and the actor engages in “interactive” dialogue.* (C) *Intermediary “break.”* (D) *The actor recites the lecture*

shown in order to create the perception of a real-time interaction (Fig. 1B), involving the actor leaning towards the camera as if the “answer call” button was just pressed followed by the actor saying “Can you hear me? (pause) Ok, then we can start.” This was followed by the actor asking the participant to recite the alphabet at their own pace for 10 s, in order to ensure that the “audio and video are correctly synced up.” The actor then shuffled papers, providing a “break” in flow (Fig. 1C), followed by a brief lecture (approximately 4 min), either about breakfast foods influencing education performance (“Breakfast grades”) or about myths and misconceptions involving monosodium glutamate (“MSG myths”) (Fig. 1D). The pre-recorded condition consisted of only this lecture segment without the interactive portion, edited to include only the portion after the papers were shuffled.

At the conclusion of the video, the experimenter re-entered the room and asked the participant to verbally respond to five questions about the lecture. All participants were asked “and you were speaking to a live person?” in addition to the question “could she/he hear and see you correctly?” if the participant was shown the “real-time” video. Participants who did not answer these questions with a “yes” response and/or did not answer at least one of the five questions correctly were excluded from analysis. If participants answered with a “no” response, they were asked to explain their answers, which were examined after the experiment and used to train the experimenters.

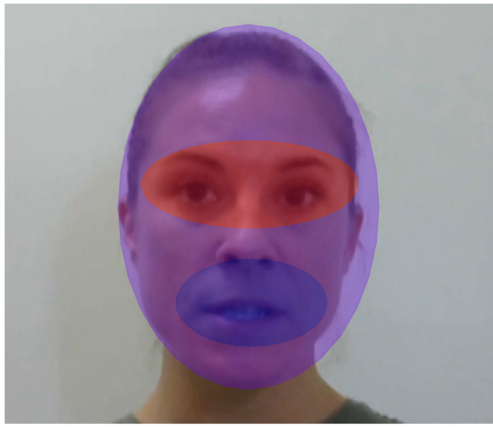
The study used a within-subjects design, with all participants engaging in both the real-time stimulus condition as well as the pre-recorded stimulus condition. Counterbalancing was used to eliminate any order effects. Out of the 98 participants used in analysis, approximately half (55) were shown the real-time video first followed by the pre-recorded video, while the remaining participants

(43) were shown the conditions in the reverse order. Additionally, approximately half the participants (50) were shown the “Breakfast grades” video followed by the “MSG myths” video, while the remaining participants (48) were shown the reverse. Both the male and female videos were shown to all participants, with one serving as the real-time and the other serving as the pre-recorded video, with the actor’s gender being counterbalanced between conditions.

### Fixation behavior analysis

Participants’ eye movements were recorded using a Tobii T60 eye-tracking system (Tobii Technology, Stockholm, Sweden), a 17-in. flat panel monitor with a screen resolution of  $1,280 \times 1,024$  pixels and a built-in infrared eye tracker with a sampling frequency of 60 Hz, chosen due to its visual similarity to a standard computer monitor. Location of gaze on the eye tracker’s screen was measured using near infrared and both bright and dark pupil-centered corneal reflection. Tracking data were analyzed with the included Tobii Studio 3.0.6 software. The eye tracker allows for head movement with a  $30 \times 22 \times 30$  cm volume when seated between 50 and 80 cm in front of the monitor, allowing for natural movement of the participant without losing tracking capabilities. All participants were tested in a quiet, well-lit room, and seated approximately 60 cm from the screen. A standardized 9-point red-on-white calibration, implemented inside the Tobii Studio software, was performed prior to data collection.

We defined three areas of interest (AOIs): the eyes, the mouth, and the whole face (Fig. 2). The whole face AOI was defined by an oval encompassing the entirety of the face. The eye AOI was defined by an oval drawn just above the



**Fig. 2** The three areas of interest (AOIs) used for fixation analyses. Full face (purple), eyes (Red) and mouth (blue)

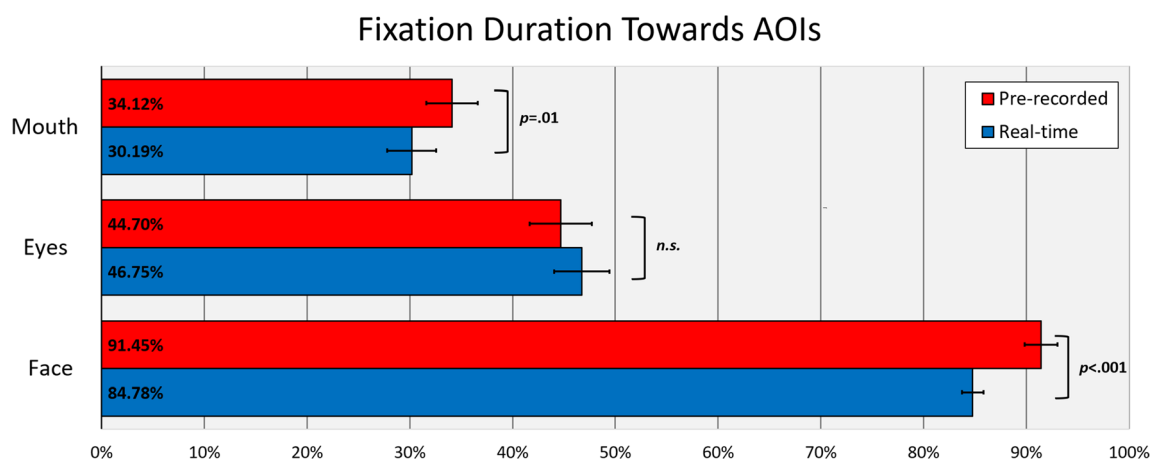
eyebrows to the bridge of the nose in the vertical axis, and from one sideburn to the other on the horizontal axis. The mouth AOI was defined by an oval drawn from the bottom of the nose to the bottom of the lips on the vertical axis, and from one cheekbone to the other on the horizontal axis. All AOIs were then enlarged by approximately 5% in all dimensions to account for small tracking errors as well as for slight head movements of the actors during the course of the video, with large head movements accounted for by moving the location of the AOIs if necessary.

Fixation behavior data points for each AOI were calculated as a ratio ranging from zero to one. Fixation data of the whole face AOI was calculated as a percentage of fixation duration occurring inside the whole face AOI divided by total running time of the video. Fixation data of the eye and mouth AOIs were calculated as a percentage of fixation duration occurring inside the eye/mouth AOI divided by fixation duration occurring inside the whole face AOI, so that eye and mouth fixation behavior could be compared across conditions even when overall face fixation behavior differed.

## Results

Seventy-five participants were excluded from analyses due to failure of deception (36), inability to answer at least one of the five questions associated with each of the lectures (24), and/or inadequate eye sampling data (15), leaving 98 participants available for analysis. Of the excluded participants for whom deception was not successful, most (73.2%) participated in the pre-recorded condition first, followed by the real-time condition. Since analysis reported no order effect, this form of counterbalancing may not be necessary in future studies.

The averaged percentage of total fixation duration was spent within the three AOIs across the two presentation conditions: real time and pre-recorded (Fig. 3). A paired-samples t-test found that participants fixated the whole-face AOI in the “real-time” condition ( $M = .85$ ,  $SD = .16$ ) significantly less than in the pre-recorded condition ( $M = .91$ ,  $SD = .10$ ),  $t(97) = -4.521$ ,  $p < .001$ ,  $BF_{01} = .001$ . Time spent fixating the mouth was also significantly greater in the pre-recorded condition ( $M = .34$ ,  $SD = .24$ ) compared to the real-time condition ( $M = .3$ ,  $SD = .25$ ),  $t(97) = -2.566$ ,  $p = .012$ ,  $BF_{01} = .53$ . However, there was no significant difference in time spent fixating the eyes between the real-time ( $M = .47$ ,  $SD = .3$ ) and the pre-recorded ( $M = .45$ ,  $SD = .27$ ) conditions,  $p = .32$ ,  $BF_{01} = 7.61$ . Comparisons between total fixation durations of the eyes versus the mouth were calculated for both the real-time condition,  $t(97) = 3.145$ ,  $p = .002$ ,  $BF_{01} = .12$ , and pre-recorded condition,  $t(97) = 2.148$ ,  $p = .034$ , with the eyes of both conditions being significantly more fixated than the mouth; however, Bayes factor analysis suggests a lack of support for this difference in the pre-recorded condition,  $BF_{01} = 1.35$ . The total number of fixations per AOI (fixation count), both overall and per visit, was found to not be significant in any context. Demographic data gathered from participants, including gender, age, cultural background, and native language, was not



**Fig. 3** Averaged percentage of total fixation duration spent within the three areas of interest across the two presentation conditions. Mouth and eye fixations were calculated as a percentage of time spent fixating the whole face

found to have influence on fixation behavior across conditions.

Performance on the five questions did not significantly differ in the real-time condition ( $M = 2.375$ ,  $SD = 1.35$ ) and the pre-recorded condition ( $M = 2.68$ ,  $SD = 1.4$ ),  $t(47) = -1.321$ ,  $p = .19$ ,  $BF_{01} = 2.47$ .

## Discussion

Overall, we found that fixation patterns of the face of a speaker during an encoding task differ when participants are aware that they are watching a pre-recorded video compared to when they believe they are interacting with a person who can see and hear them. Specifically, there was a highly significant tendency for participants engaging in a perceived real-time interaction to display greater avoidant fixation behavior, supporting the idea that social contexts draw fixations away from the face compared to when social context is not a factor. When the face was fixated, attention was directed towards the mouth for a greater percentage of time in the pre-recorded condition versus the real-time condition. This may suggest that participants are more comfortable looking directly at the mouth of a speaker when they do not believe their fixations can be observed. However, the lack of a difference in time spent fixating the eyes suggests that the additional mouth fixations in the pre-recorded condition did not come at the cost of reduced eye fixation and must have derived from reduced fixations elsewhere on the face.

Regardless of the specific mechanisms underlying the observed differences in fixation patterns, these results suggest participants were taking social/attentional considerations into account in the real-time condition. Given that encoding and memory have been found to be optimized by fixating the mouth (Vatikiotis-Bateson et al., 1998), which was reduced overall in the real-time condition, this suggests that people do not fully optimize for speech encoding alone in a live interaction. While we did not find a significant difference in performance on the follow-up questions across conditions in this study, our small set of questions was not designed or calibrated to be sensitive to potentially small differences in comprehension.

The current results add a new dimension to the extensive literature on eye fixations during speech encoding and social interactions: while it has long been known that the direction of an individual's gaze can impact the attention/fixation of an observer of that person (i.e., gaze cueing), this is the first study, to our knowledge, that demonstrates an inverse effect in which one's knowledge of being observed by another person impacts fixation behavior. The conclusions of previous studies of fixation behavior, which are based on pre-recorded videos, may need to be re-examined in light of these findings because they may not reflect the full spectrum of considerations that dictate real-world interactions. This

consideration may be particularly relevant in relation to studies of autism spectrum disorder and the gaze avoidance that is typically associated with it (Speer, Cook, McMahon, & Clark, 2007). Because social functioning tends to be compromised in autism, it is important to consider the full range of factors that may impact fixation behaviors in interpreting any observed atypicalities. The results reported here indicate an important and previously unknown factor that should be considered in future studies.

## Compliance with ethical standards

**Declaration of interest** None.

## References

- Barenholtz, E., Mavica, L., & Lewkowicz, D. J. (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition*, 147, 100–105. <https://doi.org/10.1016/j.cognition.2015.11.013>
- Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, 30(2), 285–297.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months?: The needle, the haystack, and the CHAT. *The British Journal of Psychiatry*, 161(6), 839–843.
- Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: the roles of action, competition and social content. *Quarterly Journal of Experimental Psychology* (2006), 61(7), 986–998. <https://doi.org/10.1080/17470210701410375>
- Bruner, J., & Sherwood, V. (1983). Thought, language and interaction in infancy. In J. D. Call, E. Galenson, & R. L. Tyson (Eds.), *Frontiers of infant psychiatry*. New York: Basic Books.
- Buchan, J. N., Paré, M., & Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, 2(1), 1–13. <https://doi.org/10.1080/17470910601043644>
- Buchan, J. N., Paré, M., & Munhall, K. G. (2008). The effect of varying talker identity and listening conditions on gaze behavior during audiovisual speech perception. *Brain Research*, 1242, 162–171. <https://doi.org/10.1016/j.brainres.2008.06.083>
- Burghardt, G. M., & Greene, H. W. (1990). 'Predator Simulation and Duration of Death Feigning in Neonate Hognose Snakes', *Animal Behaviour* 36(6), 1842–1843.
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. 6(5), 509–540. <https://doi.org/10.1080/135062899394920>
- Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews*, 24(6), 581–604. [https://doi.org/10.1016/S0149-7634\(00\)00025-7](https://doi.org/10.1016/S0149-7634(00)00025-7)
- Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, 99(14), 9602–9605.
- Farroni, T., Menon, E., & Johnson, M. H. (2006). Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology*, 95(4), 298–308.

- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5, 490–495.
- Glenberg, A. M., Schroeder, J. L., & Robertson, D. A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition*, 26(4), 651–658. <https://doi.org/10.3758/BF03211385>
- Gobel, M. S., Kim, H. S., & Richardson, D. C. (2015). The Dual Function of Social Gaze. *Cognition*, 136, 359–364. <https://doi.org/10.1016/j.cognition.2014.11.040>
- Griffin, Z. M., & Bock, K. (2000). What the Eyes Say About Speaking. *Psychological Science*, 11(4), 274–279. <https://doi.org/10.1111/1467-9280.00255>
- Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS ONE*, 10(8), 1–18. <https://doi.org/10.1371/journal.pone.0136905>
- Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of human evolution*, 40(5), 419–435.
- Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition*, 6, 541–567.
- Lansing, C. R., & McConkie, G. W. (2003). Word identification and eye fixation locations in visual and visual-plus-auditory presentations of spoken sentences. *Perception & Psychophysics*, 65(4), 536–552. <https://doi.org/10.3758/BF03194581>
- Lee, Y. J., Greene, H. H., Tsai, C. W., & Chou, Y. J. (2016). Differences in sequential eye movement behavior between Taiwanese and American viewers. *Frontiers in Psychology*, 7(MAY), 1–11. <https://doi.org/10.3389/fpsyg.2016.00697>
- Lusk, L. G., & Mitchel, A. D. (2016). Differential Gaze Patterns on Eyes and Mouth During Audiovisual Speech Segmentation. *Frontiers in Psychology*, 7(February), 52. <https://doi.org/10.3389/fpsyg.2016.00052>
- Maurer, D. (1985). Infants' perception of facedness. In T. M. Field & N. A. Fox (Eds.), *Social perception in infants* (pp. 73–100). Norwood, NJ: Ablex.
- Morency, L. P., Christoudias, C. M., & Darrell, T. (2006). Recognizing gaze aversion gestures in embodied conversational discourse. *Proceedings of the 8th International Conference on Multimodal Interfaces - ICMI '06*, 287. <https://doi.org/10.1145/1180995.1181051>
- Nichols, K. A., & Champness, B. G. (1971). Eye gaze and the GSR. *Journal of Experimental Social Psychology*, 7(6), 623–626.
- Scaife, M. (1976). 'The response to eye-like shapes by birds. II. The importance of staring, pairedness, and shape.' *Animal Behavior* 24, 200–206.
- Scaife, M., & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature*, 253, 265–266.
- Senju, A., Vernetti, A., Kikuchi, Y., Akechi, H., Hasegawa, T., & Johnson, M. H. (2013). Cultural background modulates how we look at other persons' gaze. *International Journal of Behavioral Development*, 37(2), 131–6. <https://doi.org/10.1177/0165025412465360>
- Speer, L. L., Cook, A. E., McMahon, W. M., & Clark, E. (2007). Face processing in children with autism: Effects of stimulus contents and type. *Autism*, 11(3), 265–277. <https://doi.org/10.1177/1362361307076925>
- Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research*, 65(4), 235–241. <https://doi.org/10.1007/s004260100059>
- Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic bulletin & review*, 9(2), 314–318.
- Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audio-visual speech perception. *Perception & Psychophysics*, 60(6), 926–940. <https://doi.org/10.3758/BF03211929>
- Võ, M. L. H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of vision*, 12(13), 3–3. <https://doi.org/10.1167/12.13.3>

**Open Practices Statement** Data and materials for the experiment reported here are available via email communication with the corresponding author

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.