



BIROn - Birkbeck Institutional Research Online

Haensel, Jen and Smith, Tim J. and Senju, Atsushi (2022) Cultural differences in mutual gaze during face-to-face interactions: a dual head-mounted eye-tracking study. *Visual Cognition* 30 (1-2), pp. 100-115. ISSN 1350-6285.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/44201/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

Cultural differences in mutual gaze during face-to-face interactions: a dual head-mounted eye-tracking study

Jennifer X. Haensel*, Tim J. Smith, and Atsushi Senju

Birkbeck, University of London
Department of Psychological Sciences
London WC1E 7HX, United Kingdom

* Corresponding author: jhaens01@mail.bbk.ac.uk

Abstract

Cross-cultural psychologists have widely discussed ‘gaze avoidance’ as a sociocultural norm to describe reduced mutual gaze in East Asians (EAs) compared to Western Caucasians (WCs). Supportive evidence is primarily based on self-reports and video recordings of face-to-face interactions, but more objective techniques that can investigate the micro-dynamics of gaze are scarce. The current study used dual head-mounted eye-tracking in EA and WC dyads to examine face looking and mutual gaze during live social interactions. Both cultural groups showed more face looking when listening than speaking, and during an introductory task compared to a storytelling game. Crucially, compared to WCs, EA dyads spent significantly more time engaging in mutual gaze, and individual instances of mutual gaze were longer in EAs for the storytelling game. Our findings challenge ‘gaze avoidance’ as a generalisable cultural observation, and highlight the need to consider contextual factors that dynamically influence gaze both within and between cultures.

Keywords: mutual gaze, head-mounted eye-tracking, culture, gaze avoidance, social interaction

Word count: 7,819 words (including abstract and captions)

Introduction

The human face allows us to identify others, infer emotional states, and participate in shared attention, highlighting the importance of visual attention to faces for successful social interactions (Bruce & Young, 1998; Haxby et al., 2000; Hoffman & Haxby, 2000). Research has consistently demonstrated an attentional bias for faces compared to other stimuli in our environment, showing that we look significantly longer at faces and also rapidly detect them within cluttered scenes (Bindemann et al., 2005; Bindemann & Lewis, 2013; Johnson et al., 1991; Langton et al., 2008; Lewis & Edmonds, 2005; Ro et al., 2001; Theeuwes & Stigchel, 2006). Such evidence is largely based on screen-based paradigms, which offer high degrees of experimental control, but typically do not allow participants to interact with the viewed face. Crucial characteristics common to ‘real-world’ social interactions – including the influence of sociocultural norms on face orienting – are therefore not taken into account, and complementary studies are needed (Cañigual & Hamilton, 2019). More recently, a greater emphasis has therefore been placed on studying social attention within a social context (Kingstone, 2009; Kingstone et al., 2008; Richardson & Gobel, 2015; Risko et al., 2012, 2016), highlighting in particular that eye gaze within more naturalistic settings not only serve to extract relevant visual information (the *encoding function*), but also signal one’s mental state to another person (the *signalling function*) (Cañigual & Hamilton, 2019; Gobel et al., 2015; Risko et al., 2016). For instance, while participants can look at face stimuli displayed on screen for as long as they wish, it would be considered socially inappropriate to gaze excessively at strangers in ‘real-world’ situations. Visual orienting to faces has been shown to decrease significantly when participants believed that a face on screen was speaking in real-time compared to a pre-recorded video (Kleiman & Barenholtz, 2020), or when they sat in the same room with another person compared to seeing a videotape of that same individual (Laidlaw et al., 2011). Passers-by close to a participant were also fixated less frequently in a ‘real-world’ setting than on screen (Foulsham et al., 2011). This suggests that the mere social presence of another person, even when not directly interacting with them, can influence the degree of face orienting.

The critical role of sociocultural norms on social attention has also been highlighted in cross-cultural investigations on mutual gaze, which we broadly define here as those periods during which two interacting individuals simultaneously gaze at each other’s *eyes* or *face* (for a discussion, see Jongerius et al., 2020). Compared to Western Caucasians, East Asian individuals are reported to engage in less mutual gaze during social interactions, which has been attributed to the sociocultural norm of ‘gaze avoidance’, with averted gaze functioning as a sign of respect in East Asian cultures (Argyle & Cook, 1976; Elzinga, 1978; Sue & Sue, 1999; Watson, 1970). In Western cultures, meanwhile, mutual gaze during social interactions is considered to positively indicate attentiveness and interest (Argyle & Cook, 1976). Using self-report measures, Argyle et al. (1986) found that British and Italian individuals rated the importance of maintaining eye contact during conversations higher than participants from Japan or Hong Kong. Content analysis furthermore revealed reduced eye contact in Japanese managers during business negotiations compared to their American counterparts (Hawrysh & Zaichkowsky, 1990). Micro-analysis of video recordings from doctor/patient conversations showed that mutual gaze in Chinese dyads was less frequent and also shorter in duration than in Canadian dyads (Li, 2004). McCarthy et al. (2006, 2008) used video recordings to code gaze directions of participants while they were answering cognitively demanding questions (e.g., solving abstract mathematical exercises), and found Japanese individuals to engage in less mutual gaze than Canadians and Trinidadians. Given the important function of eye gaze for non-verbal communication (Csibra & Gergely, 2006; Kleinke, 1986; Senju & Johnson, 2009), gaze avoidance in East Asian individuals has been widely discussed within cross-cultural research and has served as a partial account to explain observed cultural differences in social gaze behaviour more generally (e.g., Blais et al., 2008; Gobel et al., 2017). For instance, gaze avoidance could partly explain why East Asian participants fixated the eye region during screen-based face recognition tasks less than Western Caucasians (Blais et al., 2008).

While the notion of gaze avoidance can provide insights into how sociocultural norms affect eye movement behaviour during face-to-face interactions, current evidence is primarily based on self-report measures or event coding of video recordings. However, while self-reports can inform about attitudes and preferences (e.g., the importance of eye contact), ratings may not necessarily reflect an individual's actual gaze behaviour when engaged in a social interaction. Although video recordings can address this limitation, the analysis of gaze events requires the subjective judgement of multiple coders, and accurate coding is further complicated when the use of third-person perspective cameras results in insufficient resolution of the coded eye region. More recently, the advent of non-invasive infrared corneal reflection eye-tracking techniques with improved spatiotemporal accuracy and precision has provided a complementary experimental approach that can record gaze behaviour during social interactions in a more objective manner. Hessels et al. (2017) presented a two-way video setup with eye-tracking that livestreams the conversational partner's face on the participant's screen (and vice versa). With the use of half-silvered mirrors, this setup also ensured that eye contact can be achieved, unlike traditional video communication systems for which looking at the partner's eyes in the video appears as gaze avoidance due to the video's offset from the camera (see also Hessels et al., 2018). This paradigm can examine gaze behaviour to individual facial features since regions-of-interests are sufficiently large on screen, though the setup requires a physical separation between interacting individuals (for details see Hessels et al., 2017). Alternatively, wearable eye-trackers have been used to record participants' eye movements during face-to-face interactions with an experimenter (e.g., Freeth et al., 2013; Freeth & Bugembe, 2019; Haensel et al., 2020; Vabalas & Freeth, 2016). To record both individuals of a dyad, a dual head-mounted eye-tracking paradigm can be employed (e.g., Ho et al., 2015; Rogers et al., 2018). With respect to cross-cultural investigations of gaze behaviour during dyadic social interactions, only one study to date used head-mounted eye-tracking to compare face orienting and scanning strategies between British/Irish and Japanese participants (Haensel et al., 2020). Both cultural groups looked more at the face of the conversational partner (a local research assistant) during periods of listening compared to speaking. This replicated previous social interaction studies (e.g., Freeth et al., 2013; Hessels et al., 2019), suggesting a robust speech effect on face orienting. Looking at a face may reflect a way to better decode speech (Vatikiotis-Bateson et al., 1998), to signal attention to the speaker (Gobel et al., 2015; Risko et al., 2016), or an attempt to reduce cognitive load when speaking (Doherty-Sneddon & Phelps, 2005). In addition, both cultural groups also engaged in more face orienting during an introductory task relative to a storytelling game (Haensel et al., 2020), which had been suggested to result from social signalling during early introductory encounters, and the more demanding nature of the storytelling game that could have induced gaze aversion. Japanese participants also scanned the eye region more than British/Irish individuals; however, since only the eye movements of the participant – and not the research assistant – were recorded, the study cannot inform about any cultural differences in two-way gaze dynamics including gaze avoidance.

The current study

To provide empirical evidence on gaze avoidance and face orienting more generally, dual head-mounted eye-tracking techniques were adopted to compare face looking (one-way) and mutual gaze (two-way) between East Asian and Western Caucasian dyads while they introduced themselves and played a storytelling game. We defined mutual gaze here as those periods during which both participants within a dyad gazed at each other's face (see Methods for further details). Ideally, eye contact could be defined as both participants fixating the pupil/iris regions of the conversational partner; however, we examined face looks for two reasons. First, a trade-off between data quality and optimal eye-tracking equipment typically exists; for instance, while head-mounted eye-trackers with superior technical specifications allow for smaller regions-of-interests (e.g., upper face to approximate the eye region), the hardware of such models tends to obstruct the visibility of the eye region. This makes it unsuitable for dual eye-

tracking set-ups like those presented in the current study. Instead, we prioritised a headset that minimally obstructs the face region. Secondly, and critically, it should be noted that current evidence reporting gaze avoidance is based on methodologies with reduced spatiotemporal resolution compared to eye-tracking.

In line with the notion of gaze avoidance, East Asian dyads were expected to spend less time overall engaging in mutual gaze compared to Western Caucasian dyads, and were also predicted to show shorter individual instances of mutual gaze. With respect to face looking, we further expected increased face looking during listening compared to speaking periods in both cultural groups, as also reported in earlier face-to-face interaction studies (Freeth et al., 2013; Haensel et al., 2020). More face looking during the introductory period compared to the storytelling game was also expected in both cultural groups based on a previous cross-cultural social interaction study using the same experimental tasks (Haensel et al., 2020). Finally, consistent with previous face-to-face interaction studies (e.g., Vabalas & Freeth, 2016; Hessels et al., 2018), we recorded autistic traits. Given evidence pointing to higher Autism Quotient (AQ; Baron-Cohen et al., 2001) scores in East Asian compared to Western Caucasian groups (Kurita et al., 2005; Wakabayashi et al., 2006), as well as evidence suggesting differential (reduced) face orienting in autistic adults during social interactions (Freeth & Bugembe, 2019), we administered the AQ to ensure that any observed cultural differences in face looking were not simply explained by group differences in autistic traits.

Materials and Methods

Participants

Forty East Asian and 40 Western Caucasian adults were tested at Birkbeck, University of London. Each participant took part in one dyadic interaction, creating a sample of 20 dyads per cultural group. Given the challenge to conduct power analyses without effect sizes (due to a lack of studies with similar experimental paradigms), a sample size of $N = 20$ dyads per cultural group was decided in line with previous dyadic interaction studies (e.g., Ho et al., 2015). Using the effect sizes reported in Haensel et al. (2020), we find that for 90% power and an alpha-value of 0.05, a sample size of $N = 16$ per group would be required (G*Power; Faul et al., 2009). Two participants were paired if they spoke the same native language, signed up for the same experimental session, and indicated that they were unfamiliar with each other. An additional six Western Caucasian dyads and two East Asian dyads were excluded due to corneal reflection track loss of at least one participant within a dyad ($N = 7$) or misunderstanding of task instructions ($N = 1$). The two cultural groups were gender-matched, with each group consisting of 10 same-gender (5 male-male, 5 female-female) and 10 mixed-gender dyads.

Western Caucasian participants ($N = 34$ British, $N = 3$ Irish, $N = 2$ Canadian, $N = 1$ US American; $M = 26.95$ years, $SD = 9.12$ years, range = 18-53 years) were born and raised in the UK, Ireland, USA, or Canada, were of White ethnicity, had never lived in a country outside Western Europe, USA, or Canada, and indicated English as their native language. East Asian participants ($N = 16$ Japanese, $N = 24$ Chinese; $M = 26.35$ years, $SD = 7.14$ years, range = 18-55 years) were born and raised in Mainland China or Japan, were of either Chinese or Japanese ethnicity, had never lived in a country outside East Asia before coming to the UK, and indicated Mandarin or Japanese as their first language. To minimise acculturation effects, only East Asian participants who recently moved to the UK were included in the study (stay in the UK: $M = 5.1$ months, $SD = 2.2$ months, range = 5 days to 8 months). While each cultural group consisted of participants from different countries, which likely introduced a degree of cultural heterogeneity, past evidence on cultural differences between East Asians and Western Caucasians is based on individuals from countries that are included in this study (i.e., China and Japan for the East Asian group, and UK, Ireland, Canada, and US for the Western Caucasian group). All

participants had normal or corrected-to-normal vision and hearing. Participants were recruited using posters and participant databases of local universities, and study adverts were circulated in a local language school, on social networking platforms, and on community websites aimed at Japanese individuals living in the UK. The study lasted 1 hour, and each participant received £8 for their time. The study was approved by the local ethics committee of the Department of Psychological Sciences, Birkbeck, University of London, and was conducted in accordance to the Declaration of Helsinki. Each participant provided written informed consent prior to the study.

Apparatus

Eye movements were recorded using two Positive Science head-mounted eye-trackers (www.positivescience.com) at a sampling rate averaging 30Hz. Since only one adult and one infant headset were available in the lab, the infant headgear was adapted for use in adults (see Supplementary Figure 1). The headgears only differed in their physical setup, with all cameras, optics, and illuminators identical and therefore not differentially affecting data quality. The adult headgear consisted of glassless frames while the ‘infant’ headset was mounted using an elastic band. Each headset included an infrared LED, one eye camera for monocular gaze tracking, and a scene camera fitted with a wide-angle lens (field-of-view 84.28° horizontally and 69.25° vertically). Scene recordings were captured at 30fps (variable) and at 640 x 480 resolution. Each eye-tracker was connected to a MacBook that recorded and saved the data, and an additional laptop was used to transmit sound to the neighbouring room to monitor participants.

Procedure

The experimenter introduced the study and obtained written consent, and asked participants to communicate in their native language. They were informed that the content of their conversation would not be used for analysis to ensure naturalistic social interactions. Participants were also informed that the study examined cultural differences in face perception, with informal post-experiment interviews confirming that they were not aware that face orienting was investigated. The experimenter mounted the headsets on each of the two participants, who were then asked to sit at a table opposite each other at approximately 1 metre distance (see Figure 1 for a participant’s point-of-view), and fill out a demographic questionnaire. Using the livestreamed eye video, the experimenter then asked each participant to look straight ahead as well as left, right, up, and down, in order to adjust the eye camera to obtain clear images of the pupil and corneal reflection. A five-point calibration procedure was conducted independently for each participant, who was asked to fixate a small calibration object held by the experimenter at five locations in the plane of the conversational partner’s face. Calibration was performed off-line using the software Yabus (www.positivescience.com), whereby the locations of the calibration object (during fixation) was marked up in the scene video frame. Given that a live visualisation of the participant’s fixations on the calibration object was not available until the post-hoc process, accuracy or precision of calibration could not be measured during the experiment; instead, the experimenter closely monitored the participant’s eye movements via the livestreamed eye and scene video during calibration and repeated any points during which they suspected the participant had not been reliably stabilising their gaze on the calibration object. Calibration was repeated for each experimental task to protect against headgear slippage and to maximise accuracy. After successful calibration, dyads received task instructions in written form in their native language. When both participants were ready, a clapperboard was used within the dyad’s field-of-view for synchronisation of the eye-tracking data during the later analysis stage, and syncing was repeated prior to each experimental task. The experimenter then left the room to ensure the dyadic interaction was not influenced by a third person, and returned after each task for re-calibration, re-syncing, and to provide task instructions.

The experimental tasks followed a previously published paradigm (Haensel et al., 2020). The *introduction* task consisted of each participant introducing themselves for at least 30 seconds to obtain sufficient data. Participants were instructed that they could say their name, mention their occupation, or describe any hobbies, but it was made clear that they were free to talk about anything as long as they were comfortable with sharing their personal details with the conversational partner. After the introduction, dyads played two rounds of the guessing game *20 Questions*, for which one participant imagined an object while the other asked up to 20 questions to guess the object. The only permissible answers were ‘Yes’, ‘No’ or ‘I don’t know’. If a participant guessed the object correctly before reaching 10 questions, the experimenter returned and asked the dyad to play an additional round. The guessing game was included to ensure that participants felt comfortable: since *20 Questions* is a clearly structured game, it facilitates a dyadic interaction more easily than a free conversation. Face looking was not investigated for *20 Questions* since speaking individuals tend to avert their gaze away from the face during the game (Ho et al., 2015). In the *storytelling* task, each participant picked a coin from the table, looked at the year printed on the coin, and told the other participant about a personal event or experience that occurred in the given year (also for at least 30 seconds). If participants could not remember a specific event or experience, they were free to talk about one from the year before or after the coin’s date. Finally, the experimenter returned to stop the recording, and participants completed the AQ (Baron-Cohen et al., 2001).

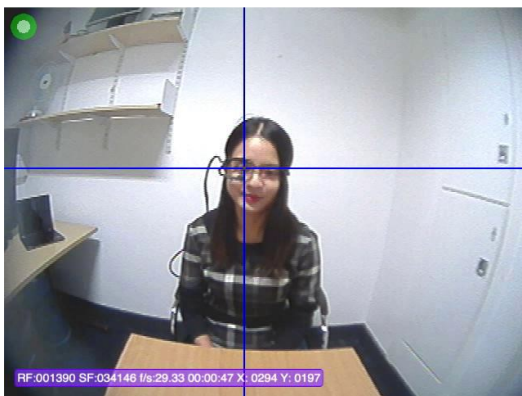


Figure 1. Participant’s point-of-view during the dyadic interaction, with gaze overlaid.

Data pre-processing

Data cleaning

Yarbus (www.positivescience.com) determines gaze coordinates by automatically tracking the pupil centre and corneal reflection in the eye video. If automatic tracking fails to accurately detect features in a frame, the bounding edges of the pupil and corneal reflection can be marked up manually. For the current study, manual selection was applied when the pupil centre or corneal reflection were positioned two or more pixels from the true centre (see Supplementary **Error! Reference source not found.**2). Given the 30Hz sampling rate, no spatial or temporal smoothing was applied, and gaze data for analysis was exported following data cleaning.

Coding of speaking and listening periods

The coding procedure for speaking and listening periods followed an existing protocol (Haensel et al., 2020). For each task, the start time of the speaking period was defined as the timing of the first frame that contained audible speech (from the speaker), whereas the end time was determined when speech ended. The start and end time of the listening period were coded accordingly in the listener’s recordings.

If a participant spoke for longer than 30 seconds, their speech was cropped to ensure that participants contributed a similar amount of data (Freeth et al., 2013). If an interruption occurred, the end time was coded as the frame preceding the interruption. A second listening/speaking period was also coded when interruptions occurred during the first 20 seconds to include more data for analysis, starting from the second sentence after the speaker resumed speaking. The start of the second sentence was selected since individuals tend to avert their gaze away from the face just after the start of speech (Ho et al., 2015). A third listening/speaking period was not used.

Regions-of-interest coding

Gaze annotations were coded semi-automatically in MATLAB (R2015a, MathWorks) following a previously published methodology for face detection and tracking (Haensel et al., 2020). Briefly, faces in scene recordings were located automatically using the Viola-Jones detector (Viola & Jones, 2001). Face detection was visualised using a rectangular bounding box (see Figure 2), and the user confirmed detection before proceeding to the next frame. When automatic detection failed, the face region was manually marked up using a rectangular box around the face according to the following guidelines: the upper and bottom edges should be along the middle of the forehead and just below the chin, respectively, and the side edges should be aligned with the sides of the face including a small margin. The Kanade-Lucas-Tomasi (KLT) algorithm (Lucas & Kanade, 1981; Tomasi & Kanade, 1991) was then adopted to track the face region using an adaptive window that changed the size, angle, and position of the bounding box. A minimum of 15 feature points was required to estimate the box, and tracking performance was visualised using a video player. Since tracking quality declines over time as a result of some feature points being lost across frames, automatic tracking terminated either when the user manually stopped the tracking process or after 150 frames – whichever came first – before returning to the initial face detection stage. For every frame, the coordinates of the four vertices of the bounding box surrounding the face were stored. Each gaze point was associated with its corresponding scene frame before classifying the point by checking whether the coordinates fell within or outside the face region. For each participant, a timeline was created with entries coded ‘1’ if the gaze point fell within the face, ‘0’ if the gaze point fell outside the face region, and ‘-1’ if an eye blink occurred (blinks were coded manually and could be distinguished from other instances of data loss since eye videos were available). An additional timeline annotated periods as listening (coded ‘0’) or speaking (coded ‘1’); this was manually coded off-line. Manual checks were performed for 20% of data (10% per cultural group), resulting in a mean accuracy of 99.56% ($SD = 0.90\%$).

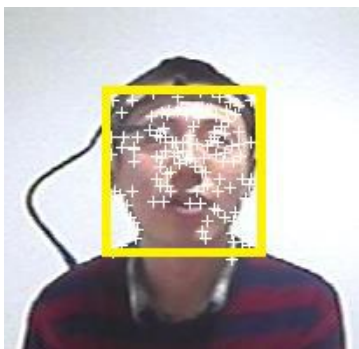


Figure 2. Bounding box surrounding the face region, with features identified for region-of-interest tracking.

Results

Three dependent variables were extracted for analysis. First, face looking was examined on a participant-level to compare Western Caucasian and East Asian individuals in their time spent gazing at the conversational partner's face across the social interaction. Secondly, we investigated mutual gaze on a dyad-level to compare the two cultural groups on the overall time spent engaging in mutual gaze over the course of the interaction. Finally, to complement the aggregated measure on mutual gaze with a more fine-grained analysis of gaze dynamics, we compared cultural groups on the durations of individual instances of mutual gaze. The analysis of each dependent variable is reported separately in the following sections.

Face looking

Each participant's time spent looking at their conversational partner's face was computed proportional to valid gaze recording time (with a cut-off at 30 seconds per task). Valid gaze recording time did not include periods of data loss (due to, e.g., blinks) in order to only consider periods for which it was possible to state whether or not participants engaged in face looking. No significant group differences were found for data loss (East Asians: $M = 17.35\%$, $SD = 8.61\%$; Western Caucasians: $M = 16.88\%$, $SD = 12.03\%$; $t(78) = 0.20$, $p = 0.843$).

A 2 (Group: Western Caucasian, East Asian) x 2 (Speech: participant is speaking, participant is listening) x 2 (Task: introduction, storytelling) mixed ANOVA was conducted. Analyses were conducted separately for listening and speaking periods, as well as for the introduction and storytelling game. The separation of analyses into these different periods was deemed necessary due to speech states and task reportedly having a major influence on gaze behaviour in dyadic interactions (e.g., Freeth et al., 2013; Hessels et al., 2019; Haensel et al., 2020). The assumption of normality was violated in some cases; significant main effects and interactions were followed up or confirmed using appropriate non-parametric tests. The analysis revealed a significant main effect of Speech ($F(1,78) = 278.50$, $p < 0.001$, $d = 2.50$; Wilcoxon Signed-Rank Test: $Z = -7.68$, $p < 0.001$, $r = 0.607$), and Task ($F(1,78) = 85.49$, $p < 0.001$, $d = 1.23$; Wilcoxon Signed-Rank Test: $Z = -6.71$, $p < 0.001$, $r = -0.530$). This suggested that participants looked more at the conversational partner's face during periods when the participants were listening compared to speaking, and during the introduction compared to the storytelling task. East Asian participants also engaged in significantly *more* face looking than Western Caucasians (Group: $F(1,78) = 5.70$, $p = 0.019$; $d = 0.14$; Mann Whitney U Test: $U = 502$, $p = 0.004$, $r = -0.321$). A significant Speech x Group interaction was found ($F(1,78) = 6.40$, $p = 0.013$, $d = 0.15$); compared to Western Caucasian participants, East Asian individuals engaged in more face looking when they were speaking ($U = 533$, $p = 0.010$, $r = -0.287$), but not when listening ($U = 610$, $p = 0.068$, with Bonferroni-corrected alpha-level of 0.025). A significant Speech x Task interaction was also found ($F(1,78) = 15.54$, $p < 0.001$, $d = 0.34$), with participants engaging in more face looking during the introduction than the storytelling task, both when speaking (Wilcoxon Signed-Rank Test: $Z = -6.40$, $p < 0.001$, $r = 0.506$) and when listening (Wilcoxon Signed-Rank Test: $Z = -4.81$, $p < 0.001$, $r = 0.380$). This reflected the Task effect, and further suggested that the difference in face looking was significantly different in the speaking condition relative to the listening condition. All other main effects and interactions were not significant (Task x Group: $F(1,78) = 3.59$, $p = 0.062$, $d = 0.09$; Speech x Task x Group: $F(1,78) = 3.28$, $p = 0.074$, $d = 0.08$)¹. Notably, when participants were listening, face looking showed ceiling effects, and variability in proportional face looking time appeared high (Figure 3; Table 1).

¹ Correlational analyses on the 40 dyads were also conducted to explore whether associations in face looking existed between individual 1 and individual 2 (from the same dyad); no significant correlations were identified (all $p > 0.05$).

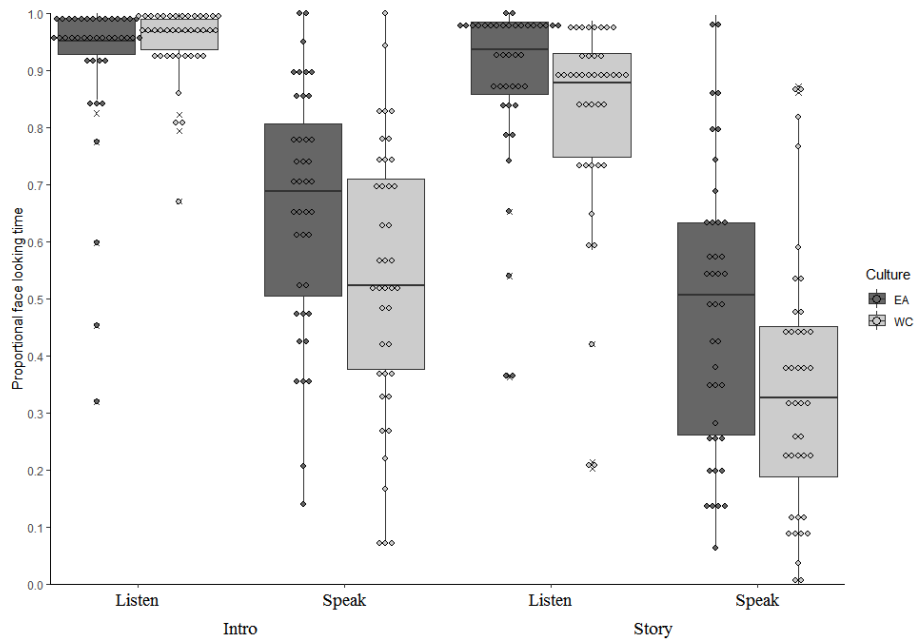


Figure 3. Proportional face looking time (range 0 to 1) of East Asian and Western Caucasian participants during the introduction and storytelling task, both when listening and when speaking (see also Table 1 for values). Gaze times are proportional to approx. 30-second recording time per condition (task/speech).

Table 1. Medians and interquartile ranges for face orienting (in %).

		East Asians <i>Mdn (IQR)</i>	Western Caucasians <i>Mdn (IQR)</i>
Introduction	Speaking	68.85 (33.69)	52.35 (36.24)
	Listening	95.19 (6.64)	96.74 (5.47)
Storytelling	Speaking	50.59 (38.57)	32.60 (31.14)
	Listening	93.69 (12.94)	87.84 (18.97)

In sum, both cultural groups looked more at the face during periods of listening than speaking, and during the introduction relative to the storytelling game. When speaking, the East Asian group also looked more at the face than Western Caucasian participants. However, face looking at the individual level does not necessarily suggest greater mutual gaze since two participants within a dyad may look at each other's face at different times. The following analysis therefore further examined mutual gaze.

Mutual gaze

Mutual gaze was defined as the periods when both participants within the same dyad simultaneously looked at each other's face, which could be determined by synchronising the event timelines of the two participants. Mutual gaze onset was considered to be the time at which both individuals within the dyad looked at the face of the other person, while mutual gaze offset was coded as the time when at least one of the participants shifted their gaze away from the face, or blinked. Mutual gaze was then calculated proportional to valid gaze recording time with a cut-off at 30 seconds per task.

A 2 (Group: Western Caucasian, East Asian) x 2 (Task: introduction, storytelling) mixed ANOVA was conducted, with all assumptions met including normality. Speech was not included as a factor given that mutual gaze depended on both participants within a dyad, and periods could not be

split into speaking and listening periods. A main effect of Task was revealed ($F(1,38) = 66.03$, $p < 0.001$, $d = 1.64$), indicating more mutual gaze during the introduction compared to the storytelling task in both cultural groups. A significant effect of Group was also found ($F(1,38) = 6.09$, $p = 0.018$, $d = 0.28$), suggesting *increased* overall mutual gaze time in East Asian relative to Western Caucasian dyads. The Task x Group interaction was not significant ($F(1,38) = 0.38$, $p = 0.541$, $d = 0.02$; Figure 4 and Table 2).

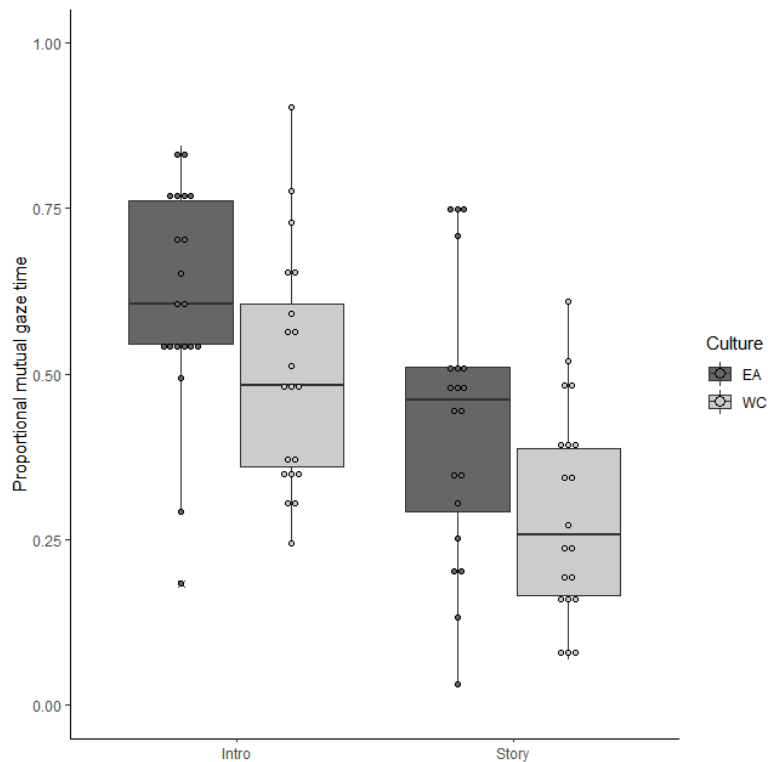


Figure 4. Proportional mutual gaze time (range 0 to 1) for East Asian and Western Caucasian participants during the introduction and the storytelling task (see also Table 2). Gaze times are proportional to approx. 1-minute recording time per task.

Table 2. Means and standard deviations for mutual face dwell time (in %).

	East Asians <i>M</i> (<i>SD</i>)	Western Caucasians <i>M</i> (<i>SD</i>)
Introduction	61.24 (16.86)	50.10 (17.78)
Storytelling	43.20 (20.79)	29.09 (15.69)

Contrary to predictions based on previous reports, East Asian dyads engaged in *more* mutual gaze than Western Caucasian dyads. It is possible, however, that individual instances of mutual gaze may have been shorter and more frequent in East Asian dyads, and longer and less frequent in Western Caucasian dyads, giving the impression of “broken” or “fleeting” periods of mutual gaze. To examine mutual gaze in a temporally-sensitive manner, the following analysis takes into account the duration of each instance of mutual gaze.

Individual mutual gaze durations

Each dyad’s durations for every mutual gaze instance was computed separately for each task (introduction and storytelling; see Supplementary Figure S3 for the distribution of durations for each dyad and task, and Supplementary Figure S4 for randomly selected scarfplots visualising mutual gaze

across the recording period). Every dyad therefore contributed numerous data points for this analysis, unlike the aggregated measures for face looking and mutual gaze (number of mutual gaze data points per dyad for East Asians, introduction: $M = 68.95$, $SD = 24.08$, range = 22–142; East Asians, storytelling, $M = 59.60$, $SD = 26.49$, range = 14–117; Western Caucasians, introduction: $M = 47.85$, $SD = 24.31$, range = 14–100; Western Caucasians, storytelling: $M = 51.75$, $SD = 24.45$, range = 12–102). Typically, individual distributions are first summarised using a central tendency measure – commonly the mean – before performing statistical analyses on the chosen summary statistic. Visual inspection of the present distributions, however, revealed heterogeneity, varying skewness, and occasionally a lack of a clear peak (see Supplementary Figure S3). Traditional analyses based on a summary statistic would therefore be misleading since no *single* value could comprehensively capture the distribution of durations. We therefore chose an alternative approach based on the analysis of quantiles, which can examine *how* distributions may differ. First, the deciles of each distribution were computed using the Harrell-Davis quantile estimator (Harrell & Davis, 1982; implemented using the *hdquantile* function from the *Hmisc* R package). This resulted in 9 values per dyad, whereby the 9 values divide the distribution into 10 parts (note that the present analysis was also conducted using quartiles and resulted in the same findings). Secondly, percentile bootstrapping (cf., Rousselet et al., 2020) was conducted using 10,000 iterations with replacement, separately for each of the 9 deciles. In each iteration, the median value of the respective decile (e.g., the first decile) was determined for the Western Caucasian group and subtracted from the median of the East Asian group. A resulting zero value would therefore indicate identical medians for the two cultural groups for the given decile, whereas a positive or negative value would point to a higher or lower median, respectively, in the East Asian compared to the Western Caucasian group. Using the values obtained from the 10,000 iterations of a given decile, a 95% confidence interval (CI) was then obtained.

For both the introduction and the storytelling game, all CIs contained the value 0 (see Supplementary Table 1), thereby indicating lack of support for any significant group differences in decile medians. In other words, we could not identify significant distributional differences between the cultural groups with respect to the duration of individual mutual gaze instances (see Figure 5 for decile medians and the bootstrapped 95% CIs).

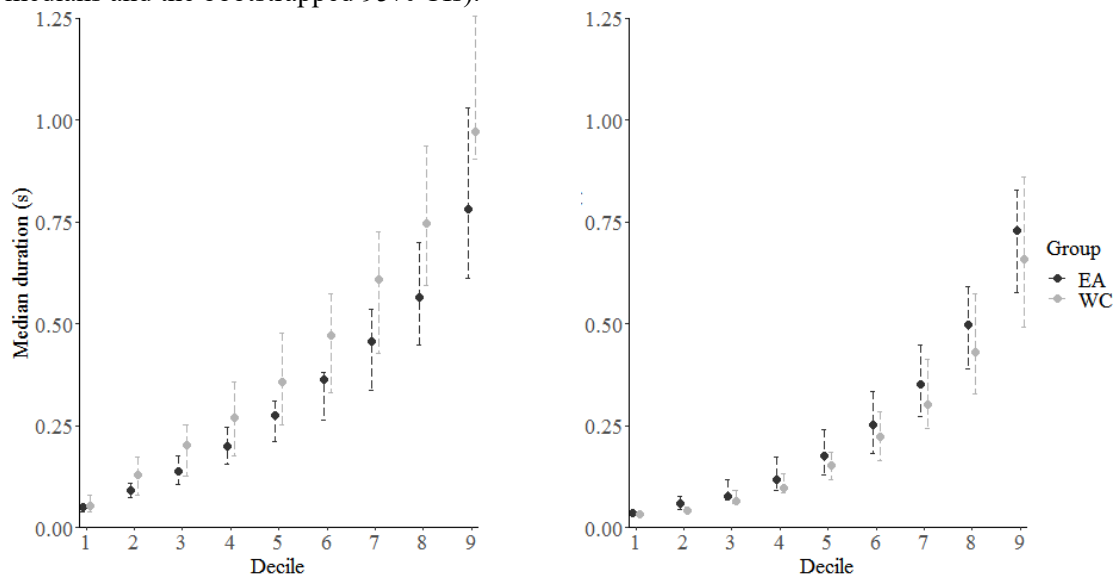


Figure 5. Decile medians of distributions for mutual gaze instances for the introduction (left) and storytelling game (right). Lines represents 95% CI.

As outlined earlier, blinks or gaze shifts away from the face by at least one member of a dyad was coded as the mutual gaze offset. We noted, however, that previous studies did not explicitly

considered blinks as event offsets. Since we were able to discern blinks from data loss using the eye videos, we repeated the analysis in an exploratory manner by only considering gaze shifts away from the face – but not blinks – as mutual gaze offsets. As expected, durations of individual mutual gaze instances increased (Figure 6). For the introduction task, the CIs for all deciles contained the value 0 as before (Supplementary Table 2), meaning that significant distributional differences between cultural groups could not be identified. For the storytelling game, however, positive CIs (greater than 0) were obtained for deciles 2 to 9. In other words, East Asian dyads showed *higher* decile medians than Western Caucasian dyads for the storytelling game Supplementary Table 2; Figure 6), pointing to *longer* mutual gaze durations in East Asian dyads.

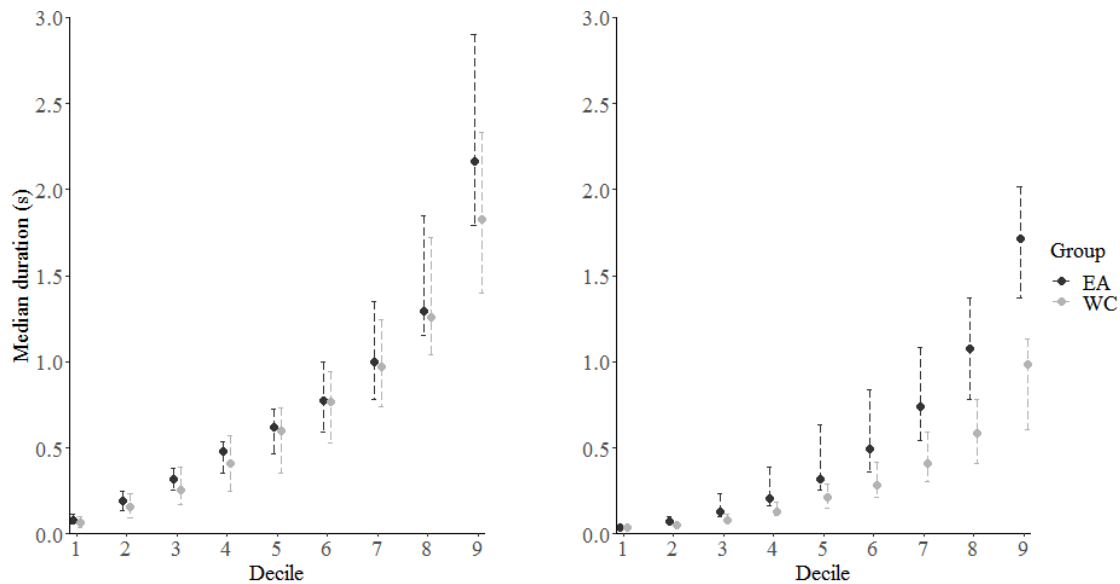


Figure 6. Decile medians of distributions for mutual gaze instances for the introduction (left) and storytelling game (right), when blinks are not considered as event offsets. Line represents 95% CI.

AQ

Western Caucasian participants obtained a significantly lower score on the AQ ($M = 14.88$, $SD = 6.97$, range = 5–36) than the East Asian group ($M = 18.05$, $SD = 5.49$, range = 7–35; $t(78) = 2.26$, $p = 0.026$, $d = 0.51$), consistent with earlier findings (Kurita et al., 2005; Wakabayashi et al., 2006). A correlational analysis investigating the relationship between each participant’s AQ score and proportional face looking did not reveal any significant correlations (all $p > 0.05$; Table 3).

Table 3. Spearman’s correlation coefficients and associated p -values for the relationship between AQ scores and face looking time during speaking and listening periods

		AQ	
		East Asian	Western Caucasian
Speaking	Face looking (Intro)	-0.059; $p = 0.718$	0.081; $p = 0.620$
	Face looking (Story)	0.088; $p = 0.590$	0.041; $p = 0.801$
Listening	Face looking (Intro)	-0.175; $p = 0.280$	-0.293; $p = 0.067$
	Face looking (Story)	0.074; $p = 0.651$	-0.134; $p = 0.410$

Discussion

The current study aimed to complement existing findings by adopting dual head-mounted eye-tracking to examine cultural differences in gaze behaviour during dyadic social interactions more reliably, objectively, and with higher spatiotemporal resolution than previous research methodologies on gaze avoidance, such as video coding or self-reports. The present findings challenge the prevailing idea of gaze avoidance in East Asian cultures, and – when taking into account previous evidence – suggest that more nuanced explanations are required when discussing cultural differences in gaze dynamics during social interactions.

In line with previous studies (e.g., Freeth et al., 2013; Haensel et al., 2020), both cultural groups showed more face looking when listening compared to speaking, indicating a robust speech effect on face gaze. Increased face looking during periods of listening could have supported participants in decoding speech (Vatikiotis-Bateson et al., 1998) as well as provided a way to signal to the conversational partner that one is still listening and attending (Gobel et al., 2015; Risko et al., 2016). Decreased face looking during periods of speaking, meanwhile, could indicate a tendency for gaze aversion to reduce cognitive load (Doherty-Sneddon & Phelps, 2005). Both cultural groups also looked more at the conversational partner's face during the introduction period compared to the storytelling game, replicating the task effect that was observed by Haensel et al. (2020). This could function as a social signal, with participants engaging in more face looking during the early stages of a novel encounter, and less face looking when they are more acquainted with each other. The more cognitively demanding storytelling game involving memory recall may have also led individuals to avert gaze (Glenberg et al., 1998), although this cannot explain the decrease in face looking for participants who were simply listening to their conversational partner's story. Face-to-face interactions are likely characterised by an interdependency of the dyadic individuals' behaviours, however, raising the possibility that the storyteller's gaze aversion may have also decreased the listening participant's face looking. With respect to cultural differences, the East Asian group engaged in more face looking than Western Caucasians, particularly during speech periods. Increased face looking in the East Asian group could indicate a greater tendency to socially signal to the conversational partner that one is still engaged in the conversation (Risko et al., 2016). Although this explanation should then also hold for periods of listening for which no cultural effect could be found, face looking when listening showed a ceiling effect in both groups, masking any potential cultural differences. Additionally, given that cultural differences in visual strategies have also been identified in a range of tasks (e.g., scene perception, Chua et al., 2005; visual search, Ueda et al., 2018), future studies could benefit from exploring the relationship between these culture-typical patterns of visual perception and gaze behaviour in social communicative contexts.

With respect to overall time spent engaging in mutual gaze, dyads gazed at each other more during the introductory task compared to the storytelling game. Since individuals (on a participant-level) showed more face looking during the introduction, this may have provided more opportunities for mutual gaze (on a dyad-level) and could further reflect greater two-way social signalling during earlier stages of meeting a person. Crucially, East Asian dyads exhibited more mutual gaze than Western Caucasian dyads, challenging the prevailing notion of gaze avoidance in East Asian cultures. Existing findings in support of gaze avoidance – based on more subjective measures such as self-reports administered in each cultural group (Akechi et al., 2013; Argyle et al., 1986) – may not fully explain the cultural differences in overt gaze behaviour during social interactions. Furthermore, current evidence based on video recordings may have been specific to interactions requiring abstract thinking (McCarthy et al., 2006, 2008), with gaze behaviour during more naturalistic dyadic interactions manifesting differently. Our findings question gaze avoidance in East Asian cultures as a generalisable cultural observation, and we put forward the need to consider the dynamic nature of social interactions (see also Hessels, 2020, for a discussion). Contextual factors such as speech or task demands can

modulate the degree to which mutual gaze is observed and may further influence gaze behaviour differently across cultural groups. Situational aspects could therefore influence cultural observations for gaze avoidance; for instance, while individuals with an East Asian cultural background may gaze less at passers-by than Western Caucasians, it could be more socially appropriate to look at a person when engaged in a dyadic social interaction. Yet again, we suggest that culture-specific norms can interact with other contextual factors; for example, social status and hierarchy, i.e. factors that were not examined in the present study, may play a greater role in East Asian communities compared to Western cultures (cf. Gobel et al., 2015, 2017), which could reduce face orienting to an individual of higher status. Furthermore, while we broadly compared participants from East Asian and Western cultures in this study, it should be noted that each cultural group was not entirely homogeneous, and it is possible that culture-specific norms impacted participants within the same cultural group differently. The two groups may have also differed in their degree of heterogeneity; for instance, most participants in the Western Caucasian group were British, whereas the East Asian group consisted of both Chinese and Japanese individuals. Although we conducted checks to ensure that face looking and mutual gaze for did not significantly differ when comparing Chinese and Japanese participants in the present study, or when comparing groups based on other factors such as gender or AQ (see Supplementary Materials), future cross-cultural studies should be acutely aware of the possibility of intragroup heterogeneity.

The analysis on the durations of individual mutual gaze instances further revealed that East Asian dyads showing *higher* decile medians than Western Caucasians – again going into the opposite direction of the expected gaze avoidance effect. Notably, this was observed only when blinks were not considered a break in mutual gaze. Given that blinks were coded as a separate gaze event and each frame was manually corrected if automatic tracking failed, the present findings cannot be solely attributed to errors in blink coding as a result of, e.g., flicker. Removing blink instances as event offsets increased individual mutual gaze durations, as expected, and our findings highlight the effect of absence/presence of blinks on temporally-sensitive analyses. Indeed, the observed mutual gaze durations in this study appeared much shorter compared to previous reports (e.g., Argyle et al., 1986; Binetti et al., 2016). Binetti et al. (2016), for instance, found a preferred mean mutual gaze duration of 3.3 seconds when participants were asked to indicate their level of (dis)comfort while maintaining eye contact with a dynamic face identity displayed on a screen. Unlike such screen-based paradigms, participants in the present paradigm were engaged in a live dyadic interaction and the social presence of the conversational partner may have influenced mutual gaze behaviour. This idea is supported by findings from a dyadic interaction study (conducted in a Western culture) reporting eye contact to last for an average of 360 milliseconds when blinks did not serve as event offsets (Rogers et al., 2018). Crucially for this study, blink behaviour did not give rise to gaze *avoidance* in East Asian dyads, but we suggest future studies to consider the possibility that blinks could serve as a way to break mutual gaze in a subtle, socially appropriate manner, without a gaze shift away from the face. Although eye blinks are sometimes considered as artefacts that function as a biological mechanism (e.g., to protect the corneal surface; Ousler et al., 2008), blinks have also been linked to cognitive processes (Hirokawa et al., 2004; Holland & Tarlow, 1975) and have been shown to be relevant in social interactions, with longer blinks more likely to occur during mutual gaze and serving as social signals to indicate understanding (Hömke et al., 2017) – highlighting the influence of blink coding on the analysis of social gaze.

Dual head-mounted eye-tracking can provide detailed insight into gaze behaviour during dyadic social interactions, but naturally comes with some limitations that need to be acknowledged. The eye-tracking equipment – and the eye camera arm in particular – obstructed parts of the face being viewed (Figure 1). The Positive Science system was chosen in this study since the headsets would not obstruct the face as substantially as some ‘goggle’-type models – a crucial property given that the study looked at face looking during a dyadic interaction – but could still have served as a visual distraction and

thereby affected gaze behaviour. Importantly, however, we did not expect any systematic differences between cultural groups that could have led to the observed results, and the present study defined mutual gaze as periods during which individuals looked the conversational partner’s face, and not the eyes specifically. Furthermore, while head-mounted eye-tracking data is more spatially sensitive than coding of gaze direction in video recordings, spatial precision and accuracy are unlikely to be sufficient for highly detailed regions-of-interest coding of the individual eyes. It is thus possible that a period was flagged as mutual gaze when one participant was looking at the mouth and the other at the eyes. We generated spatially sensitive, descriptive gaze difference maps for face scanning (see Haensel et al., 2020). Briefly, all gaze points that fell within the face region were re-mapped into a unified coordinate space that represented the face region, and gaze points were collapsed across time to produce heatmaps. For both tasks, East Asian participants appeared to scan the eye region more, while Western Caucasians exhibited greater gaze distributions around the eyes and across the face (Figure 7). While these visualisations cannot conclude for certain that mutual gaze in the East Asian group was directed at the eye region, the heatmaps corroborate the present findings in a descriptive manner; for instance, East Asian participants did not appear to spend a large proportion of time looking at the mouth.

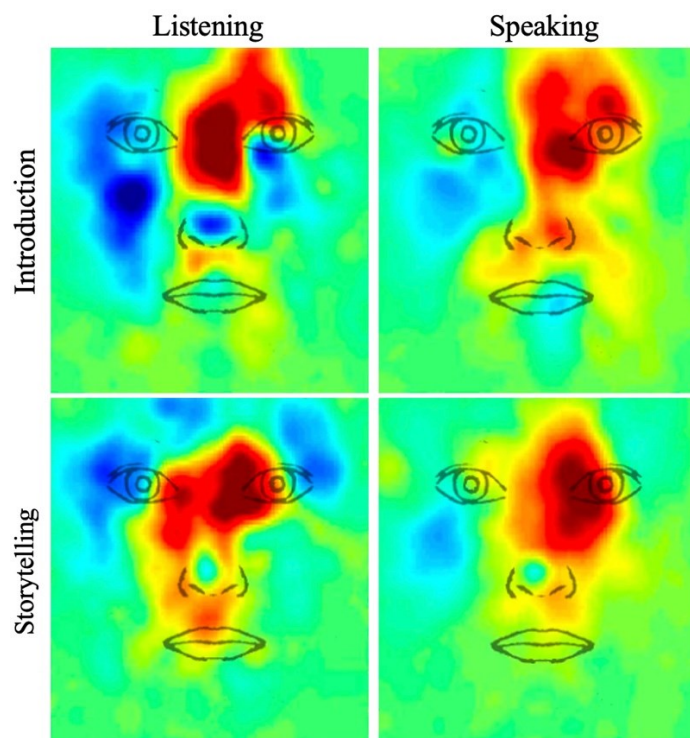


Figure 7. Descriptive heat maps visualising cultural differences in face scanning during periods of listening (left) and speaking (right) for the *introduction* task and *storytelling game*. Red and blue colours depict regions that East Asians and Western Caucasians scanned more, respectively.

While some studies have subdivided the face into smaller regions-of-interest (e.g., upper versus lower face to approximate eye and mouth regions), this was done using different head-mounted eye-tracking models. For example, participants in the cross-cultural social interaction study by Haensel et al. (2020) wore SMI glasses; while these glasses obstruct the visibility of the eye region for the observer, they provide higher pixel resolution and a closer shot of the face, and therefore superior technical specifications for more detailed regions-of-interest coding. Certain head movements of either the participant or their conversational partner can also result in different sizing of the upper and lower face region; for instance, for a head that is tilted downward, the lower face will appear smaller than the upper face on video. While this is not an issue for dyadic interaction studies that involve a confederate, who

can be trained to keep their head still (Haensel et al., 2020), providing such instructions to participants in dual eye-tracking set-ups could make the interaction less ‘naturalistic’. Consequently, a trade-off will typically exist when setting up a social interaction paradigm involving eye-tracking; for example, while a chin- and headrest can improve data quality, this would interfere with a dyadic interaction, and while dual eye-tracking set-ups such as those introduced by Hessels et al. (2017) provide an intriguing approach for examining scanning behaviour to individual facial features during live dyadic interactions, this may not be suitable for every study given that faces are presented on a screen. Researchers therefore should evaluate the pros and cons of their experimental set-up and equipment to identify a methodology that best fits the given research question (see also Valtakari et al., 2021). Critically, although future research will benefit from more spatially sensitive measures for ‘real-world’ social interaction paradigms, current evidence reporting gaze avoidance is not based on more spatially precise techniques than those employed here. Our findings challenge the notion of gaze avoidance, consistent with earlier evidence that East Asian individuals tend to fixate the conversational partner’s eye region more than Western Caucasians (Haensel et al., 2020).

In sum, the present study used dual head-mounted eye-tracking techniques to complement existing findings on cultural differences in mutual gaze. We suggest that the precise manifestations of cultural differences in social gaze is modulated by contextual factors, highlighting also the need to consider the dynamic nature of dyadic interactions – also beyond cross-cultural research.

Acknowledgements

This work was supported by the Medical Research Council (MR/K016806/1, G1100252) and Wellcome Trust/Birkbeck Institutional Strategic Support Fund (204770/Z/16/Z). We thank Pakling An and Nanami Harada for help with translations.

Disclosure statement

The authors declare no competing interests.

Data availability statement

The raw datasets (scene videos) are not publicly available since they disclose personally identifiable information. However, all processed data analysed for this study are made available in anonymised form with DOI 10.17605/OSF.IO/A7U6D (<https://osf.io/a7u6d/>).

References

- Akechi, H., Senju, A., Uibo, H., Kikuchi, Y., Hasegawa, T., & Hietanen, J. K. (2013). Attention to Eye Contact in the West and East: Autonomic Responses and Evaluative Ratings. *PLoS ONE*, 8(3), e59312. <https://doi.org/10.1371/journal.pone.0059312>
- Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge University Press.
- Argyle, M., Henderson, M., Bond, M., Iizuka, Y., & Contarello, A. (1986). Cross-Cultural Variations in Relationship Rules. *International Journal of Psychology*, 21(1–4), 287–315. <https://doi.org/10.1080/00207598608247591>
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and

- Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, 31(1), 5–17. <https://doi.org/10.1023/A:1005653411471>
- Bindemann, M., Burton, A. M., Hooge, I. T. C., Jenkins, R., & de Haan, E. H. F. (2005). Faces retain attention. *Psychonomic Bulletin & Review*, 12(6), 1048–1053. <https://doi.org/10.3758/BF03206442>
- Bindemann, M., & Lewis, M. B. (2013). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin & Review*, 20(6), 1140–1145. <https://doi.org/10.3758/s13423-013-0445-9>
- Binetti, N., Harrison, C., Coutrot, A., Johnston, A., & Mareschal, I. (2016). Pupil dilation as an index of preferred mutual gaze duration. *Royal Society Open Science*, 3(7), 160086. <https://doi.org/10.1098/rsos.160086>
- Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture Shapes How We Look at Faces. *PLoS ONE*, 3(8). <https://doi.org/10.1371/journal.pone.0003022>
- Bruce, V., & Young, A. (1998). *In the eye of the beholder: The science of face perception*. Oxford University Press.
- Cañigueral, R., & Hamilton, A. F. de C. (2019). The Role of Eye Gaze During Natural Social Interactions in Typical and Autistic People. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.00560>
- Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 12629–12633. <https://doi.org/10.1073/pnas.0506162102>
- Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development* (pp. 249–274). Oxford University Press.
- Doherty-Sneddon, G., & Phelps, F. G. (2005). Gaze aversion: A response to cognitive or social difficulty? *Memory & Cognition*, 33(4), 727–733. <https://doi.org/10.3758/BF03195338>
- Elzinga, R. H. (1978). Temporal Organization of Conversation. *Sociolinguistics Newsletter*, 9(2), 29–31.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. [doi:10.3758/BRM.41.4.1149](https://doi.org/10.3758/BRM.41.4.1149)
- Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, 51(17), 1920–1931. <https://doi.org/10.1016/j.visres.2011.07.002>
- Freeth, M., & Bugembe, P. (2019). Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults? *Autism*, 23(2), 503–513. <https://doi.org/10.1177/1362361318756786>

- Freeth, M., Foulsham, T., & Kingstone, A. (2013). What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. *PLoS ONE*, 8(1), e53286. <https://doi.org/10.1371/journal.pone.0053286>
- 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., Chen, A., & Richardson, D. C. (2017). How different cultures look at faces depends on the interpersonal context. *Canadian Journal of Experimental Psychology*, 71(3), 258–264. <https://doi.org/10.1037/cep0000119>
- 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>
- Haensel, J. X., Danvers, M., Ishikawa, M., Itakura, S., Tucciarelli, R., Smith, T. J., & Senju, A. (2020). Culture modulates face scanning during dyadic social interactions. *Scientific Reports*, 10(1), 1–11. <https://doi.org/10.1038/s41598-020-58802-0>
- Harrell, F. E., & Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika*, 69(3), 635–640. <https://doi.org/10.1093/biomet/69.3.635>
- Hawrysh, B. M., & Zaichkowsky, J. L. (1990). Cultural Approaches to Negotiations: Understanding the Japanese. *International Marketing Review*, 7(2). <https://doi.org/10.1108/EUM0000000001530>
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223–233. [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)
- Hessels, R. S. (2020). How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-020-01715-w>
- Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2017). Gaze behavior to faces during dyadic interaction. *Canadian Journal of Experimental Psychology*, 71(3), 226–242. <https://doi.org/10.1037/cep0000113>
- Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T. C., & Kemner, C. (2018). Eye contact takes two – autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology*, 9(2), 1–17. <https://doi.org/10.5127/jep.062917>
- Hessels, R. S., Holleman, G. A., Kingstone, A., Hooge, I. T., & Kemner, C. (2019). Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors. *Cognition*, 184, 28–43. <https://doi.org/10.1016/j.cognition.2018.12.005>
- Hirokawa, K., Yagi, A., & Miyata, Y. (2004). Comparison of Blinking Behavior during Listening to and Speaking in Japanese and English. *Perceptual and Motor Skills*, 98(2), 463–472. <https://doi.org/10.2466/pms.98.2.463-472>

- Ho, S., Foulsham, T., & Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PLoS ONE*, *10*(8), e0136905. <https://doi.org/10.1371/journal.pone.0136905>
- Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84. <https://doi.org/10.1038/71152>
- Holland, M. K., & Tarlow, G. (1975). Blinking and Thinking. *Perceptual and Motor Skills*, *41*(2), 403–406. <https://doi.org/10.2466/pms.1975.41.2.403>
- Hömke, P., Holler, J., & Levinson, S. C. (2017). Eye Blinking as Addressee Feedback in Face-To-Face Conversation. *Research on Language and Social Interaction*, *50*(1), 54–70. <https://doi.org/10.1080/08351813.2017.1262143>
- Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1), 1–19. [https://doi.org/10.1016/0010-0277\(91\)90045-6](https://doi.org/10.1016/0010-0277(91)90045-6)
- Jongerius, C., Hessels, R. S., Romijn, J. A., Smets, E. M. A., & Hillen, M. A. (2020). The Measurement of Eye Contact in Human Interactions: A Scoping Review. *Journal of Nonverbal Behavior*, *44*(3), 363–389. <https://doi.org/10.1007/s10919-020-00333-3>
- Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology*, *19*(1), 52–56. <https://doi.org/10.1016/j.conb.2009.05.004>
- Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive Ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340. <https://doi.org/10.1348/000712607X251243>
- Kleiman, M. J., & Barenholtz, E. (2020). Perception of being observed by a speaker alters gaze behavior. *Attention, Perception, & Psychophysics*, *82*(5), 2195–2200. <https://doi.org/10.3758/s13414-020-01981-9>
- Klinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1), 78–100. <https://doi.org/10.1037/0033-2909.100.1.78>
- Kurita, H., Koyama, T., & Osada, H. (2005). Autism-Spectrum Quotient–Japanese version and its short forms for screening normally intelligent persons with pervasive developmental disorders. *Psychiatry and Clinical Neurosciences*, *59*(4), 490–496. <https://doi.org/10.1111/j.1440-1819.2005.01403.x>
- Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences*, *108*(14), 5548–5553. <https://doi.org/10.1073/pnas.1017022108>
- Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, *107*(1), 330–342. <https://doi.org/10.1016/j.cognition.2007.07.012>

- Lewis, M., & Edmonds, A. (2005). Searching for faces in scrambled scenes. *Visual Cognition*, *12*(7), 1309–1336. <https://doi.org/10.1080/13506280444000535>
- Li, H. Z. (2004). Culture and Gaze Direction in Conversation. *RASK*, *20*, 3–26.
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 674–679.
- McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2006). Cultural Display Rules Drive Eye Gaze During Thinking. *Journal of Cross-Cultural Psychology*, *37*(6), 717–722. <https://doi.org/10.1177/0022022106292079>
- McCarthy, A., Lee, K., Itakura, S., & Muir, D. W. (2008). Gaze Display When Thinking Depends on Culture and Context. *Journal of Cross-Cultural Psychology*, *39*(6), 716–729. <https://doi.org/10.1177/0022022108323807>
- Ousler, G. W., Hagberg, K. W., Schindelar, M., Welch, D., & Abelson, M. B. (2008). The Ocular Protection Index. *Cornea*, *27*, 509–513. <https://doi.org/10.1097/ICO.0b013e31816583f6>.
- Richardson, D. C., & Gobel, M. S. (2015). Social Attention. In J. Fawcett, E. Risko, & A. Kingstone (Eds.), *Handbook of Attention* (pp. 46–61). MIT Press.
- Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: Toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, *6*, 1–11. <https://doi.org/10.3389/fnhum.2012.00143>
- Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. *Current Directions in Psychological Science*, *25*(1), 70–74. <https://doi.org/10.1177/0963721415617806>
- Ro, T., Russell, C., & Lavie, N. (2001). Changing Faces: A Detection Advantage in the Flicker Paradigm. *Psychological Science*, *12*(1), 94–99. <https://doi.org/10.1111/1467-9280.00317>
- Rogers, S. L., Speelman, C. P., Guidetti, O., & Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific Reports*, *8*(1), 4271. <https://doi.org/10.1038/s41598-018-22726-7>
- Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2020). The percentile bootstrap: A primer with step-by-step instructions in R. *Advances in Methods and Practices in Psychological Science*.
- Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, *13*(3), 127–134. <https://doi.org/10.1016/j.tics.2008.11.009>
- Sue, D. W., & Sue, D. (1999). *Counseling the culturally different: Theory and practice* (3rd ed.). John Wiley & Sons Inc.
- Theeuwes, J., & Stigchel, S. V. der. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition*, *13*(6), 657–665. <https://doi.org/10.1080/13506280500410949>
- Tomasí, C., & Kanade, T. (1991). *Detection and tracking of point features*.

- Ueda, Y., Chen, L., Kopecky, J., Cramer, E. S., Rensink, R. A., Meyer, D. E., Kitayama, S., & Saiki, J. (2018). Cultural differences in visual search for geometric figures. *Cognitive Science*, *42*(1), 286-310. <https://doi.org/10.1111/cogs.12490>
- Vabalas, A., & Freeth, M. (2016). Brief Report: Patterns of Eye Movements in Face to Face Conversation are Associated with Autistic Traits: Evidence from a Student Sample. *Journal of Autism and Developmental Disorders*, *46*(1), 305–314. <https://doi.org/10.1007/s10803-015-2546-y>
- Valtakari, N. V., Hooge, I. T., Viktorsson, C., Nyström, P., Falck-Ytter, T., & Hessels, R. S. (2021). Eye tracking in human interaction: Possibilities and limitations. *Behavior Research Methods*, 1-17. <https://doi.org/10.3758/s13428-020-01517-x>
- Vatikiotis-Bateson, E., Eigsti, I.-M., Yano, S., & Munhall, K. G. (1998). Eye movement of perceivers during audiovisual speech perception. *Perception & Psychophysics*, *60*(6), 926–940. <https://doi.org/10.3758/BF03211929>
- Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 1*, I–I. <https://doi.org/10.1109/CVPR.2001.990517>
- Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., & Tojo, Y. (2006). The Autism-Spectrum Quotient (AQ) in Japan: A Cross-Cultural Comparison. *Journal of Autism and Developmental Disorders*, *36*(2), 263–270. <https://doi.org/10.1007/s10803-005-0061-2>
- Watson, O. M. (1970). Proxemic Behavior: A Cross-Cultural Study. In *Proxemic Behavior*. De Gruyter Mouton.