Behav Res
DOI 10.3758/s13428-017-0913-7

Online webcam-based eye tracking in cognitive science: A first look

Kilian Semmelmann[1] · **Sarah Weigelt**[1]

Abstract Online experimentation is emerging in many areas of cognitive psychology as a viable alternative or supplement to classical in-lab experimentation. While performance- and reaction-time-based paradigms are covered in recent studies, one instrument of cognitive psychology has not received much attention up to now: eye tracking. In this study, we used JavaScript-based eye tracking algorithms recently made available by Papoutsaki et al. (*International Joint Conference on Artificial Intelligence*, 2016) together with consumer-grade webcams to investigate the potential of online eye tracking to benefit from the common advantages of online data conduction. We compared three in-lab conducted tasks (fixation, pursuit, and free viewing) with online-acquired data to analyze the spatial precision in the first two, and replicability of well-known gazing patterns in the third task. Our results indicate that in-lab data exhibit an offset of about 172 px (15% of screen size, 3.94° visual angle) in the fixation task, while online data is slightly less accurate (18% of screen size, 207 px), and shows higher variance. The same results were found for the pursuit task with a constant offset during the stimulus movement (211 px in-lab, 216 px online). In the free-viewing task, we were able to replicate the high attention attribution to eyes (28.25%) compared to other key regions like the nose (9.71%) and mouth (4.00%). Overall, we found web technology-based eye tracking to be suitable for all three tasks and are confident that the required hard- and software will be improved continuously for even more sophisticated experimental paradigms in all of cognitive psychology.

Keywords Online experiment · Web technology · Eye tracking · Online study · Cognitive psychology

Eye tracking provides a unique way to observe the allocation of human attention in an extrinsic manner. By identifying where a person looks, scientists are able to identify what guides human visual attention. Yarbus (1967) was one of the first to quantify this attentional value by recording eye-gaze patterns, and since then techniques to measure and algorithms to interpret eye movements have only been improved. The main approaches to utilize eye-tracking data are to measure the onset of a saccade – a jerk-like re-allocation of foveal fixation – and smooth pursuit movements, the stable tracking of moving objects. While the former introduces a shift in human attention by voluntarily or involuntarily centering a point of interest at the physically highest resolution (the fovea), pursuit movements are used to follow motion. Therefore, using eye fixations as a metric of information gathering reveals when, for how long, and how often someone looks at certain parts of an image to obtain visual information.

Consequently, eye tracking can be found in many areas of science (see Duchowski, 2002, for an overview). From basic research on perception (e.g., Chua, Boland, & Nisbett, 2005) or memory (e.g., Allopenna, Magnuson, & Tanenhaus, 1998), over more applied fields in marketing (e.g., Wedel & Pieters, 2000), industrial engineering (e.g., Chapman, Underwood, & Roberts, 2002), learning (van Gog & Scheiter, 2010) to clinical settings, e.g., to quantify differences between special populations and controls (e.g., Boraston & Blakemore, 2007; Holzman, Proctor, & Hughes, 1973). Taken together, the ease

✉ Kilian Semmelmann
kilian.semmelmann@rub.de

[1] Developmental Neuropsychology, Department of Psychology, Ruhr-University Bochum, Universitätsstr. 150, Bochum 44801, Germany

Published online: 07 June 2017

of application makes eye tracking one of the widest spread scientific tools.

On the other hand, eye tracking can be considered a laborious approach. It relies on expensive and stationary in-lab equipment, usually a meticulous calibration procedure as well as the need for a scientist to perform the experiment. Fortunately, recently a movement in the eye-tracking research community has emerged that tries to move away from the classical costly setup towards cheaper alternatives of eye-tracking devices (Dalmaijer, 2014), open source software for data analyses (Dalmaijer, Mathôt, & Van der Stigchel, 2013; Mathôt, Schreij, & Theeuwes, 2012), wearable eye trackers (Bulling & Gellersen, 2010), or even the use of webcams for recording gaze patterns (Valenti, Staiano, Sebe, & Gevers, 2009). Together, these endeavors indicate that scientists in the field of eye tracking wish to broaden its applications by making use of new technologies as supplements or even substitutes to classical tracking setups. Beyond a reduction of costs, the exchange of open source software and new areas of data acquisition (e.g., mobile recordings through wearables or large-scale longitudinal studies through consumer-grade devices) can be defined as main factors for a change of behavior in the community.

As such, the movement of the eye-tracking research community is comparable to a similar movement in the area of psychophysics. Following the success of online-conducted questionnaires (Birnbaum, 2000, amongst others), several studies furthered methodological investigations of online data acquisition towards error-rate based studies (Germine et al., 2012), and lately on comparing millisecond-based reaction time studies (Semmelmann & Weigelt, 2016). The general findings are that an overwhelming majority of experimental effects can be reproduced in an online setting and – following earlier literature about online participants (Gosling, Vazire, Srivastava, & John, 2000) – the data are of the same quality as in-lab acquired data. The reasons for moving towards online data conduction are numerous: Main points include the opportunity to obtain large, more diverse participant samples required for the generalization of results, a higher speed of data acquisition thanks to parallel, autonomous data recordings, independence of available soft- and hardware, and lower costs due to less time involvement of participants as well as scientists. Obviously, if the eye-tracking community could utilize these advantages in their movement towards new areas, likewise larger and more diverse participant samples and the use of new technology could be facilitated to a greater extent.

Until recently, however, it was not possible to efficiently combine eye tracking and online research due to the lack of eye-tracking libraries implemented in JavaScript, the most commonly used script language in online research. But that changed with two publications by Xu et al. (2015) and Papoutsaki et al. (2016). Xu and colleagues concentrated on generating saliency maps of natural images and reported an astoundingly low 30-pixel median error of intended and recorded eye position compared to an Eyelink 1000 (1,000 Hz), when using an external webcam and a head rest in an in-lab environment. Papoutsaki et al. on the other hand did not rely on an explicit calibration but used a self-training model that used mouse movements and clicks as fixations. Following the correlation between mouse position and visual attention (Chen, 2001), they used clicks and mouse movements to constantly re-train the model with gaze estimation. Their approach achieved between about 170- and 260-pixel error for the webcam approach compared to a 50-Hz Tobii EyeX eye tracker. While these studies concentrated on massive data sets under restrained circumstances (Xu et al., 2015) or on-the-fly feedback that relies on user interaction (Papoutsaki et al., 2016), we found a lack of information for a more applied direction in cognitive psychology that resembles classical in-lab studies, with a focus on fast, low-cost data acquisition and easy access to a broader population.

In this study, we employed three paradigms to investigate the general accuracy and applicability of a JavaScript-based webcam eye-tracking implementation. The first was a fixation task, in which a dot was shown and the participant was asked to fixate the dot, similar to a calibration procedure. The second task was a pursuit task (e.g., Fukushima, Fukushima, Warabi, & Barnes, 2013), in which participants were instructed to follow the movement of the target stimulus. In the third task, an image of a face was shown and observers were free to look at whatever caught their attention (similar to Dalmaijer, 2014), thereby creating an attentional map of the image. This task was meant to complement the rather technical nature of the first two tasks by introducing a semantic layer of attention and trying to reproduce earlier studies about the different fixation patterns in relevant areas of faces (eyes vs. mouth and nose in Westerners; Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Caldara, Zhou, & Miellet, 2010).

While we analyzed each of the tasks in a classical in-lab setting (also using consumer-grade webcams) to have a baseline of accuracy, we extended our data set by acquiring data through an online (crowdscience) environment. This allowed taking a step further from the technical nature of hard- and software accuracy in a controlled environment to changing the conduction surrounding and identifying potential differences between the technology itself and effects of online data acquisition. While we generally assumed that online data might be noisier and potentially less accurate than in-lab acquired data, we specifically investigated the differences, and discuss potentials solutions for emerging offsets. Taken together, we see this as one of the first investigations on the potential and possible limitations of webcam-based online eye tracking in the field of cognitive psychology and beyond.

Methods

Hardware and software

In our in-lab setting, participants were seated 50 cm from the 15-in. display of a MacBook Pro with a resolution of 1,680 × 1,050 px and its built-in webcam (720p FaceTime HD). The experiment itself was programmed in HTML5, supported by jQuery 1.12.3 and php 5.3.3 on an Apache 2.6.18. The eye-tracking algorithms were taken from Papoutsaki et al. (2016; https://webgazer.cs.brown.edu) due to the good documentation and ease of integration into a new application and were modified to fit our purpose (available at the Open Science Framework; https://osf.io/jmz79). Mainly, we removed the reliance on mouse movements and clicks and integrated a calibration-based model training method. Despite not being intended for external calibration, we were able to get the eye-tracking algorithm running "out of the box" and to adapt it to our needs through the clear object-oriented and well-documented build.

Stimuli and design

Participants completed six blocks, two of each task (fixation, pursuit, free viewing), in random order. Before each block, the participant was shown instructions on how to position him/herself (see Fig. 1), a live feed of his/her webcam, and a short notice to try to stop moving the head once the block starts.

Calibration and validation Each of the blocks started with a calibration phase, in which the participant was instructed to simply fixate a dot. In each calibration step a 50 × 50 px red-colored (#dc4a47) dot appeared and – after a 750-ms delay – we started to collect 20 samples, one every 100 ms, to train the ridge regression (for details of the implementation, see Papoutsaki et al., 2016). In total, 26 calibration dots appeared per calibration phase, equiprobable over 13 calibration positions (see Fig. 2 for details), adding up to 520 calibration samples per participant per block.

After the calibration, validation started. In each validation phase, five points appeared one after the other, in random order, on each of the four edge positions of the calibration (positions being 20% top, bottom, left, and right) and in the center. Each validation sample required the fixation to be within 200 px (4.58° in-lab) of the center of the dot. If a position was not successfully validated within 20 s, the participant was sent back to the instructional page and had to re-calibrate. If there were five unsuccessful attempts in a row, the user was informed that his hard-/software or setup was not suitable for the experiment and the experiment ended automatically. If the webcam was successfully calibrated and validated, the actual task started. Before each task, the participant received a written on-screen instruction.

Tasks Each trial of the fixation task started with a 1,500 ms black fixation cross, followed by a 500-ms blank screen, before a 50 × 50 px blue-colored (RGB: 221, 73, 75) dot appeared for 2,000 ms. The fixation cross and dots each covered 50 × 50 px, which corresponded to 1 cm on-screen size in the in-lab setting, translating into 1.15° of visual angle.

In the pursuit task a ramp stimulus (Fukushima et al., 2013) was used. At the beginning of each trial, a black dot appeared at one of four starting positions, each having 20% margin towards either the top or bottom and either left or right (see Fig. 2). The dot turned red after 1,500 ms and moved towards another position within 2,000 ms. Starting position and motion were randomly picked from the 12 possible combinations. The distance covered was 596 px (11.5 cm, 13.12°) in the vertical, 1,007 px (20 cm, 22.62°) in the horizontal, and 1,170 px (23 cm, 25.91°) in the diagonal motions.

Lastly, the free-viewing task consisted of a centered 1,000 ms fixation cross, followed by a 500-ms blank screen, before one of 30 facial images (15 female, 15 male) was shown for 3,000 ms. The image was aligned with its center at 25% from either the left border or the center of the screen and was scaled to fill 100% of the height of the screen. Each task ended with a blank screen, before the next trial started. On average, the faces covered 14 × 16 cm (615 × 820 px) of the in-lab screen, corresponding to 15.94° x 18.18° of visual angle. All images were taken from the Glasgow Unfamiliar Face Database (Burton,

Fig. 1 Instructions to the participant in order to achieve high data quality

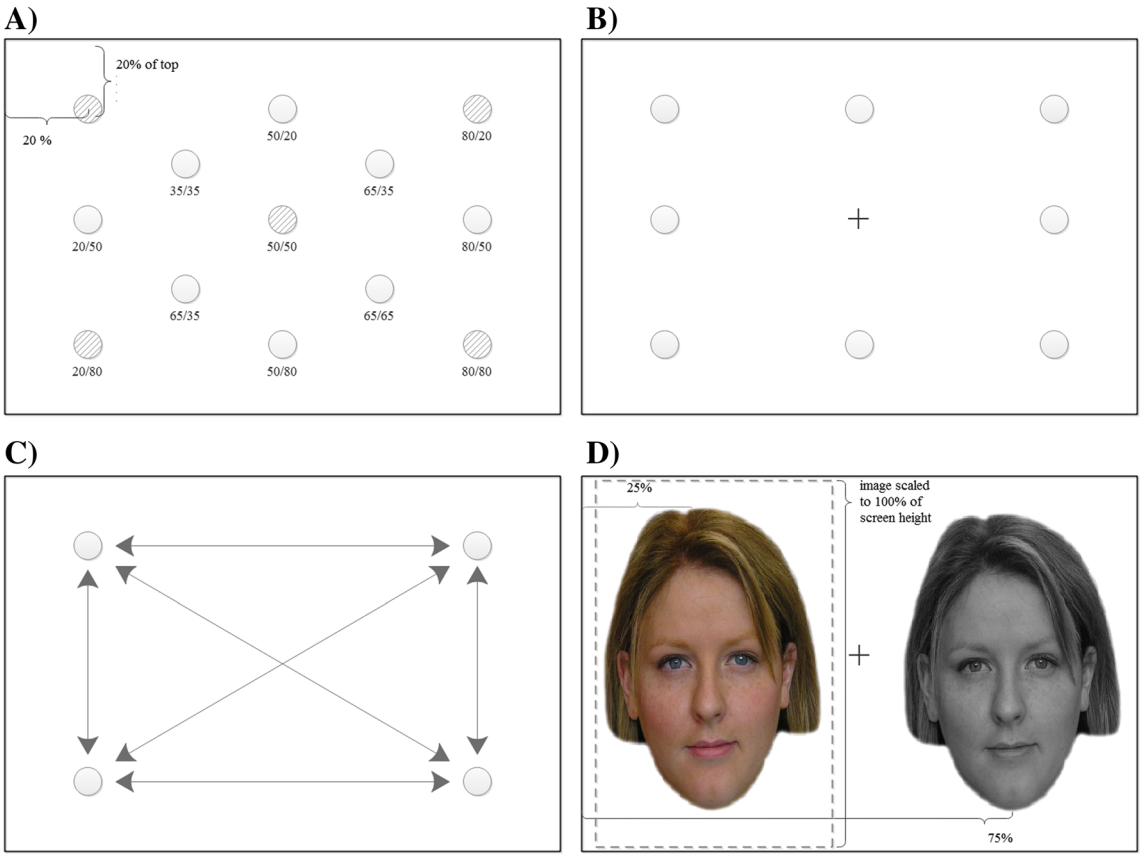


Fig. 2 Arrangement of stimuli. (**A**) The calibration and validation phase. A calibration dot could appear at any of the 13 positions, while validation dots only appeared at positions marked with a diagonal pattern. (**B**) The fixation task positions. (**C**) The pursuit task positions. In the pursuit task, movement could be vertical, horizontal, or diagonal, therefore creating 12 possible trial types. The last task, free viewing (**D**), showed a face image centered either at 25% or 75% of screen width, while being scaled to 100% of screen height

White, & McNeill, 2010; http://www.facevar.com/downloads/gufd), cropped to the faces with hair only and presented on a grey background.

Participants

The experiment was approved by the local ethics committee and participants provided a completed consent form before being able to continue on the website. All data was analyzed on the fly on the participants' computers (either on the MacBooks for the in-lab or their personal systems in the online case), thus only sending the eye position to our server. Compared to sending whole video files, this approach allows increasing anonymity and decreasing data size to a large degree.

Thirty in-lab participants (age M = 22.68 years, SD = 3.71, range 19–32 (one did not report his age), 76% female, 24% male) were recruited at the Ruhr-Universität Bochum and participated for course credit or 5€ for the 30-min experiment. For the crowdscience approach, we used www.crowdflower.com to distribute the experiment. In total, we had 84 consent form transmissions that divided into 52 incomplete and 32 complete data sets. Of the incomplete ones, 27 started the first, 11 the second, four the third, and three the fourth calibration before quitting the experiment. Overall there were 59 failed validation trials over 25 participants of the incomplete data sets, 33 over 32 participants in the complete online data sets, and four over 30 participants in the in-lab case. These numbers show that most dropouts occur due to either (a) privacy concerns or (b) high perceived effort to complete the task before starting the calibration. Only complete online participations were rewarded with a US$4 compensation. We distributed the experiment in Germany only to avoid potential cultural influences in comparison to the in-lab participants. The mean age was 39.92 years (SD = 12.88, range 20–62; two did not report their age), 29% identified as female, and 71% as male. On average, in-lab participants indicated fewer visual impairments (17% glasses, 14% corrected through contact lenses, 7% uncorrected, vs. 62% without visual impairment) than online participants (32% glasses, 11% corrected through contact lenses, 46% good vision, 11% uncorrected). Overall, one in-lab participant had to be excluded because of missing data (transmission error) and four crowdflower data

sets were removed due to multiple participation[1]), and conducting the task without visual aids despite having themselves identified as visually impaired. This left us with 29 data sets for the in-lab and 28 for the online setting for all further analyses.

Analysis

For the fixation task As an initial indication of whether data quality was high enough to detect changes in fixation position through saccades, we plotted the fixation positions over time for the fixation task. This allowed observing the change from the middle of the screen (fixation cross) towards the appearing dot at a very basic level, given a sufficiently accurate algorithm. To account for saccadic preparation and the saccade itself (Walker, Walker, Husain, & Kennard, 2000), we assumed that the participants' fixations should have arrived at the target point 1,000 ms after the target appears (3,000–4,000 ms). During this time frame, we further quantified the accuracy of the approach by calculating the offset as the Euclidean distance between intended target position and recorded gaze estimation. Accordingly, analyses of variance (ANOVAs) quantified a potential difference between position of the target and experimental setting (in-lab vs. online).

We followed a similar approach with the pursuit task. Here, we expected a saccade towards the initial fixation point (appears at trials start), followed by a steady Euclidean distance to the target that should not significantly change if participants followed the movement (starting at 1,500 ms) and the algorithm was sensitive enough to follow the variability in fixation position. Consequently, we first calculated the offset per participant at between 2,000–4,000 ms and calculated a t-test to identify the potential difference between in-lab and online data acquisition. Beyond this main analysis, to gain a better understanding of the overall data quality, we calculated a principle component analysis (PCA) for each setting and for each direction of movement to be able to compare the intended target movement to the recorded viewing behavior. Furthermore, to compare our results to earlier pursuit literature (e.g., Duchowski, 2002; Lisberger, Morris, & Tychsen, 1987), we calculated pursuit speed.

As we instructed participants to freely look at the stimulus as they like in the free-viewing task, we initially used a very basic differentiation of fixated display side to check whether an appropriate saccade towards the target stimulus was detectable, or participants did not look at the presented image at all. For more in-depth analyses, we defined regions of interest (ROIs) in each image shown. This was done by dividing each image into a grid with 5% distances and then marking the coordinates (similar to Dalmaijer, 2014; Pelphrey et al., 2002) for image (full image), head (leaving out white space on the image, on average 71.32% of the image area), inner face (not including forehead, chin or ears, 26.99%), and the key regions of the face, namely eyes (without eye brows, 7.98%), nose (2.72%), and mouth (3.63%), as can be seen in Fig. 2. These ROI analyses allowed checking for a replication of distribution of fixation time on key facial features (e.g., Caldara et al., 2010) and potential differences in experimental setting (in-lab vs. online).

Besides those main factors, we further analyzed potential non-task-specific factors: Here, as a first factor, we compared overall experimental duration to reveal whether online participants take longer or shorter and/or take more or less pauses compared to in-lab observers. Additionally, we assumed a difference in available hardware power between in-lab and home systems. Thus, we calculated the framerate per second (fps) our application was to achieve and investigated potential influences on the results above. These factors are informative whether potential accuracy differences are due to behavioral (longer pauses, less concentration) or computational (low-performing computer systems) reasons.

Please note that all analyses were intentionally carried out on unprocessed, unsmoothed data without outlier removal to enable a picture of the raw data quality and to identify potential noise. Additional processing (which often depends on the research question) obviously would allow improving the accuracy of results in subsequent studies. To allow detailed questions on how data quality improvements can enhance the analyses to be answered, we published the data online for interested colleagues to apply their methodology (see above). Furthermore, due to the differences in screen size some statistical tests were calculated based on percentage of screen size instead of absolute pixels. This is necessary as (a) the position of stimuli was calculated on screen size (e.g., "20% from the left edge") and (b) only through that way do we get an equal comparability over the different screen sizes and resolutions used in the online case. We tried to note both pixel-wise (for a clearer mental representation) and percentual values (for better comparability), whenever possible.

[1] We used the service of www.crowdflower.com in our study, as the more common Amazon Turk (AMT) was not accessible in Germany at the time. Furthermore, crowdflower accesses many different external mini-task job sites and distributes their jobs to them, potentially allowing a better generalization of the obtained data compared to AMT, where studies have questioned the diversity of participants (Stewart et al., 2015). While we had great experiences in psychophysical studies with crowdflower, we encountered one severe case of fraud during this specific work. Especially one person used the success code, necessary to pass the system's compensation program, multiple times on multiple computers with different accounts, thereby circumventing the system and breaching the terms of service. Despite the literature being mostly in favor of fair participation of online studies (e.g., Gosling, Vazire, Srivastava, & John, 2000), and our overall positive experience with online participants, we highly recommend that multiple participation does not go unchecked. Methods to do so include checking IP address, user agent, screen size, or simply the use of a success-based reimbursement system that only will compensate users after successful participation and data validation through the scientists ("bonus" payment).

Results

Fixation task

By plotting the mean fixation position over time (Fig. 3), we analyzed the first pattern of viewing behavior through the spatial changes over time; most notably we expected a jump towards target on stimulus onset (saccade) and an improved offset afterwards. The saccade from the fixation point towards the target between 2,000 and about 2,600 ms was clearly identified, as shown in a change of mean Euclidean distance of real target position and eye fixation from 493 px (pre-stimulus, 500–1,500 ms) to 189 px (stimulus present, 3,000–4,000 ms). On average, the saccade started at 2,375 ms and took 450 ms until it fully reached the target in the in-lab, and started at 2,250 ms and took 750 ms in the online case. This high duration might be attributed to slight offsets in saccade start either due to individual differences or due to performance offsets (please refer to the Appendix for a visualizing single subject graph).

Taking screen size into account (Fig. 4), a repeated-measures ANOVA indicated no significant difference for the between-subject-factor experimental setting (in-lab vs. online), $F(1, 55) = 3.77$, $p = .06$, but for the Greenhouse-Geissner corrected ($\varepsilon = 0.65$) within-subject-factor position, $F(4.55, 250.25) = 24.40$, $p < .001$. Holm-corrected post-hoc t-tests revealed that positions at left and right bottom of the screen had a higher offset than middle or top positions (for detailed results see Appendix). Additionally, we observed a significantly higher variance in the online compared to the in-lab setting (post-hoc F-test on position offset variance in screen size % between 3,000 and 4,000 ms, $F(28, 27) = 0.25$, $p < .001$). To verify the non-significance of the between-subject factor setting, we calculated the Bayes Factor, $BF_{10} = 1.259$, which argues for insufficient data to reach a decisive conclusion. Overall, in-lab data showed an average Euclidean offset of 15% ($SD = 4$) of screen size, while online data had 18% ($SD = 9$) when considering the data between 3,000 and 4,000 ms.

Additionally, to differentiate between random and systematic offset, we calculated the general average offset (compared to Euclidean distance we used before) of samples for both settings. This approach cancels out random noise (i.e., samples that gather around the target position in any direction) by averaging over fixation samples and yields a more systematic variation from target position. We found −69 px (6.1%) and −54 px (5.5%) offset for the in-lab and online cases, respectively, which did not differ significantly for the between-subject factor setting, $t(43.85) = 0.06$, $p = .95$. Please see the Appendix for a table of these offsets over target positions.

In short, while both experimental settings showed a clearly detectable saccade of about 300 px (22% of screen size, 7.39° visual angle in the in-lab case) towards the target position, in-lab conducted data, especially in the upper parts of the screen, seemed to be more accurate due to shorter saccadic duration and lower variance (with 171 px, 15%, 3.94° visual angle offset for in-lab, 207 px, 18% online).

Pursuit task

Investigating the pursuit movement task (Fig. 5), we focus on investigating the temporal-spatial pattern of fixations: If eye position can be found to move towards fixation and kept at a constant offset afterwards, this means that the task is done

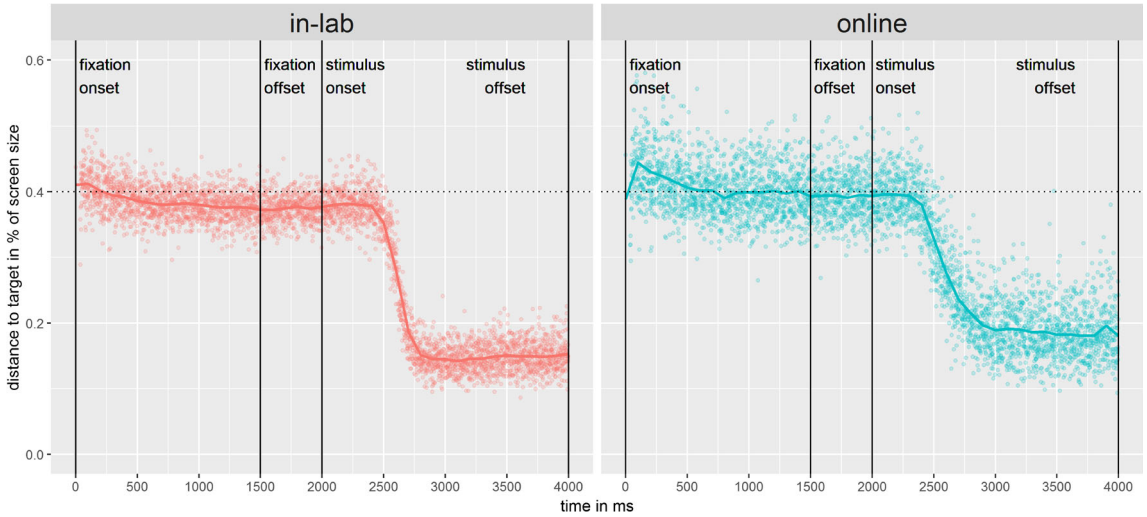


Fig. 3 Fixation task results. Each dot denotes a single recorded data point in distance to target in percentage of screen size over time. A saccade is clearly visible after target onset (2,000 ms) in both experimental settings (red in-lab setting and cyan online setting). On average, the saccade took 450 ms (750 ms in the online case) and showed an accuracy of 171 px (207 px online), which translates to about 3.94° visual angle in the in-lab setting

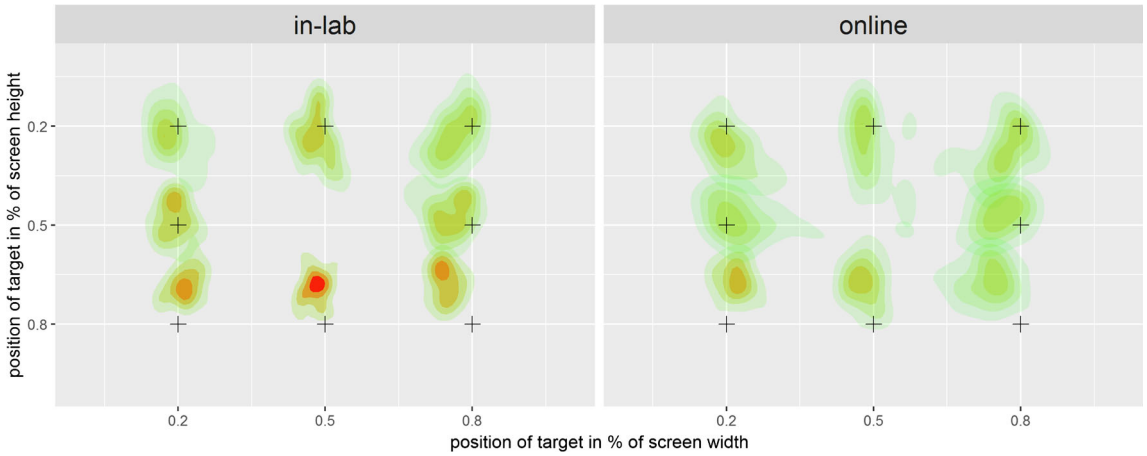


Fig. 4 Heatmap for the fixation task. Each black cross shows the position of a fixation target. The heatmaps (observed gaze behavior) show a closer match to the intended target position (cross) for targets on the top of the screen. Additionally, the higher variance in the online (right) in comparison to the in-lab data (left) can be observed

correctly by the participants and the method is accurate enough to detect the motion of the stimulus. We found a clear saccade towards the point of fixation before the movement started (375–875 ms, 49% to 18% screen size, 642 px to 211 px decrease, which equates to a 9.85° saccade in visual angle for in-lab and a 375–1,125 ms, 50% to 20%, 612 px to 216 px decrease for the online case). During the movement of the target (after 2,000 ms), the offset was stable until the point reached the target position, where the trial ended. This was confirmed through a regression of 10-ms binned averaged data between 2,000 and 4,000 ms, R^2 = .005, $F(1, 167)$ = 0.87, p = .35. With regard to potential differences between in-lab and online data acquisition accuracy, a Welch's t-test on the offset in % screen size did not find a difference during the motion part of the task, averaged between 2,000 and 4,000 ms, $t(55)$ = 1.63, p = .11. A subsequent Bayes Factor analysis argued for too few samples to reach conclusive results in this regard, BF_{10} = 0.802. Descriptively, in-lab data seem to capture an increase in the offset at the start of the movement, which declines until the end of the trial, resembling a catch-up saccade (e.g., Fukushima et al., 2013).

With regard to pursuit speed, we calculated the difference between 10-ms binned time points in percentage of screen size and found the peak at 570 ms for the in-lab and at 510 ms for the online case, shortly after the onset of the fixation dot. The average maximal change of 18.20%/10 ms (238 px/10 ms)

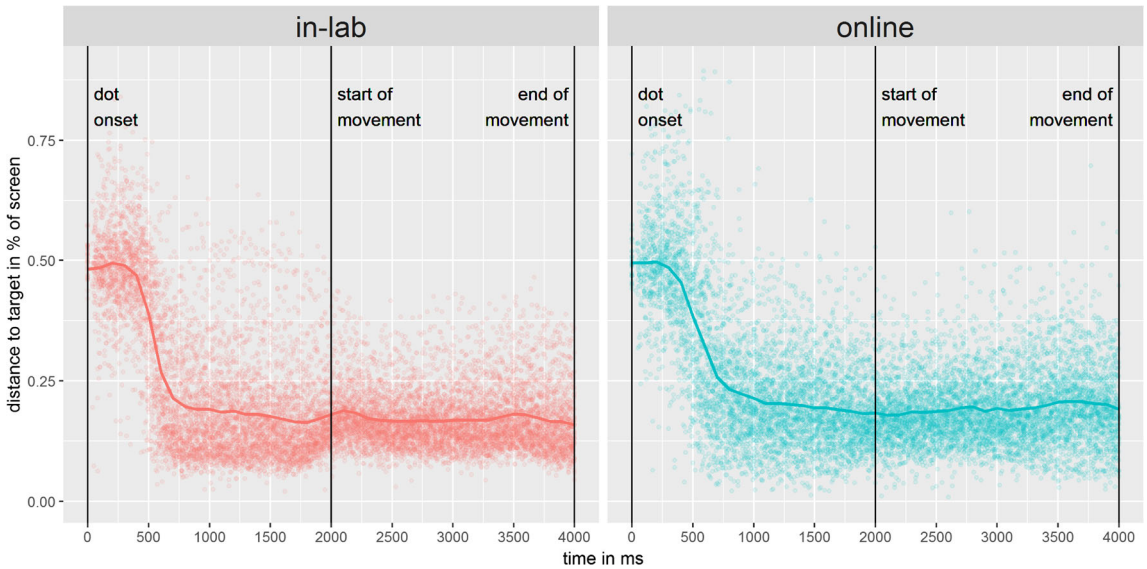


Fig. 5 Pursuit task results. Each dot represents one gaze estimation offset in percentage of screen size over time in the pursuit task. The initial saccade towards the fixation dot starts at about 375 ms, takes 625 ms on average to fully reach target position, and reduces the offset from 50% to 19% (627 to 214 px). This offset is held constant in both experimental settings once the movement starts, yet a catch-up saccade can be observed in the in-lab data (red), but not in the online data (cyan)

translates to a speed of 65.90°/s for the in-lab case. For the online case, the maximal change was 16.93%/10 ms (210 px/10 ms). Speed decreases until around 2,000 ms and stays at 8.80%, 103 px (37.89°/s) in the in-lab, and 10.47%, 118 px in the online condition (Fig. 6) during the motion.

Additionally, to have a better general understanding on how well the movement paths were represented by the singular fixation samples, we fit PCAs on each start-end combination (Fig. 7). A PCA is necessary to account for the multidimensional data (x, y coordinates) that takes into account both values, compared to linear regression, especially in the case of vertical eye movements. While most of the data fit the real movement well, we can observe impaired fixation recordings for the horizontal conditions on the lower part with a consistent offset towards the top of the screen, while the directional movement was well aligned (horizontal lines on graphs 1B, 1D, 2B, 2D in Fig. 7). In sum, we were able to identify the initial saccade to the starting point, a constant offset during the motion, but found a potentially lower spatial resolution on the lower part of the screen, regardless of setting.

Free viewing

For the free-viewing data, participants, who had less than 50% fixation samples on the image were excluded (N = 1). Fig. 8 shows the initial differentiation, i.e., the identification of the participants' gaze towards the sides of the screen where the target was shown. To quantify a potential difference between experimental settings, we calculated the mean difference in gaze position between 3,000 and 4,500 ms for each participant per screen side. A repeated measures ANOVA indicated no significant difference for the between-factor setting, $F(1, 54) = 1.17$, $p = .15$, but for the within factor target side, $F(1, 54) = 1287.33$, $p < .001$ and a significant interaction between those factors, $F(1, 54) = 7.46$, $p = .009$, based on the fact that the difference between left and right is greater in the in-lab condition (45.39%) than in the online condition (38.98%), $t(47.34) = 2.70$, $p = .01$.

To compare fixation samples in relevant areas of the image, we plotted the percentage of samples per region of interest (ROI; Fig. 9) between 3,000 and 4,000 ms in Fig. 10. Through this analysis we can discern the amount of attention attribution to each area of the overall face, therefore finding areas of higher interest to the observer. For the in-lab data, 30.07% ($SD = 12.17$) of fixations were recorded at the eyes, 11.01% ($SD = 5.65$) at the nose, and 4.43% ($SD = 3.01$) at the mouth. The values for the online data were 26.28% ($SD = 10.68$) eyes, 8.31% ($SD = 5.00$) nose, and 3.87% ($SD = 3.14$) mouth (for details see Fig. 10 and the Appendix). An ANOVA on all on-image ROIs indicated a significant difference for the between-factor setting, $F(1, 54) = 8.33$, $p < .001$, and for the Greenhouse-Geissner corrected ($\varepsilon = 0.49$) within-factor ROI, $F(2.5, 132.3) = 161.68$, $p < .001$. As expected, Holm-corrected pairwise t-tests revealed a higher amount of fixation samples on the eye region than nose or mouth. All t-tests and their corresponding significance values are shown in the supplemental files. To supplement these results, we calculated the Bayes Factor and found a difference for the within-factor ROI, $BF_{10} = 9.821e^{66}$, but could not reach a conclusive result for the between-factor setting, $BF_{10} = 0.274$. In sum, we found evidence that a replication of typical gaze patterns towards faces is possible in-lab and online, yet online data seem to be slightly less accurate than in-lab acquired data.

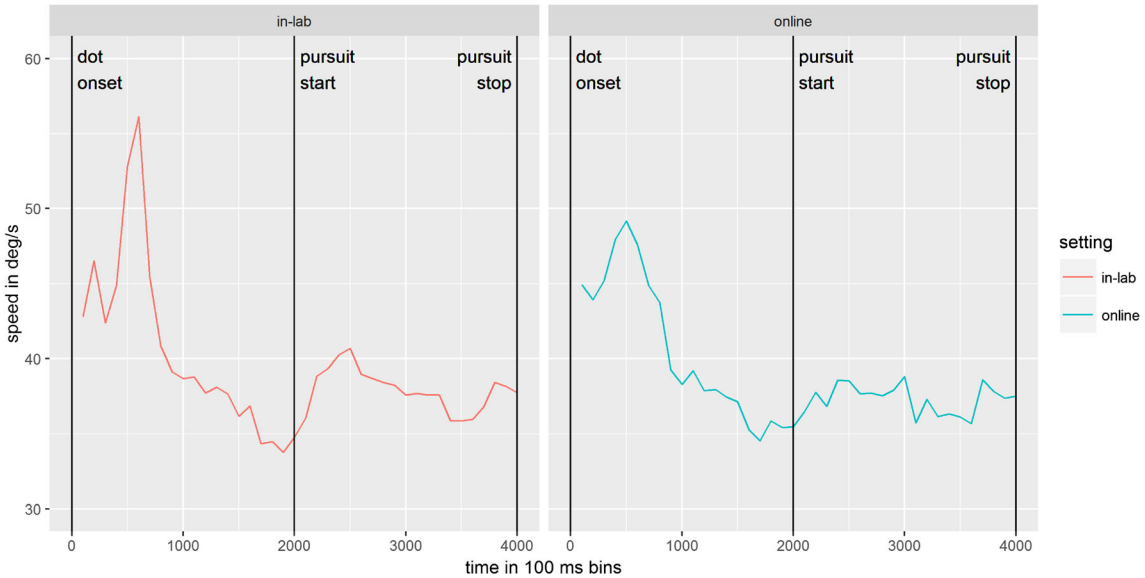


Fig. 6 Speed profile of the pursuit task. Speed was calculated from the Euclidean distance between two gaze estimations and transformed into visual angle (based on 50-cm viewing distance in the in-lab case). The highest speed can be observed shortly after onset of the fixation dot, declines until the movement starts, and then is held steady over the course of the movement

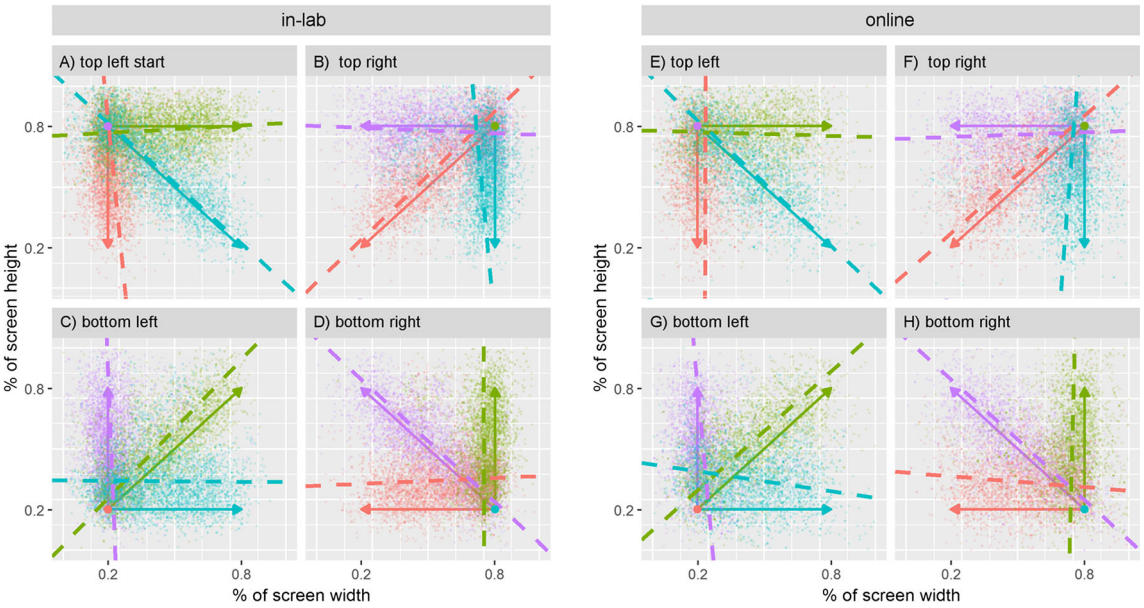


Fig. 7 Pursuit-fitted principle component analysis (PCA). Pursuit movement data are split into starting position and are color coded for end positions. Each arrow denotes the target movement path, while the dashed lines are the PCA-fitted lines. While most PCAs fit the data well, in graphs C, D, G, and H a considerable offset towards the top can be found in all the horizontal movements on the lower portion of the screen, while the motion direction is kept accurately

Additional analyses

As additional factors to consider between in-lab and online conducted data, we investigated the experimental duration and the frame rate per second (fps). The experimental duration was significantly higher for online (M = 43.54 minutes, SD = 29.98, range 25.38–169.76) than for in-lab (M = 28.06 minutes, SD = 29.98, range 22.60–37.31) participants, as indicated by a Welch t-test, $t(27.48) = 2.72$, $p = .01$. With regard to frame rate (fps), a t-test indicated a significantly higher fps (Fig. 11) for in-lab (M = 18.71, SD = 1.44, range 15.39–21.44) data acquisition than for those at home (M = 14.04, SD = 6.68, range 4.50–25.69), $t(29.41) = 3.62$, $p = .001$. In short, online conducted data exhibits a lower sampling rate and participants take more time to finish the whole experiment compared to in-lab acquisition.

Discussion

Using consumer-grade webcams our aim was to establish a first common ground of the potential and limitations of

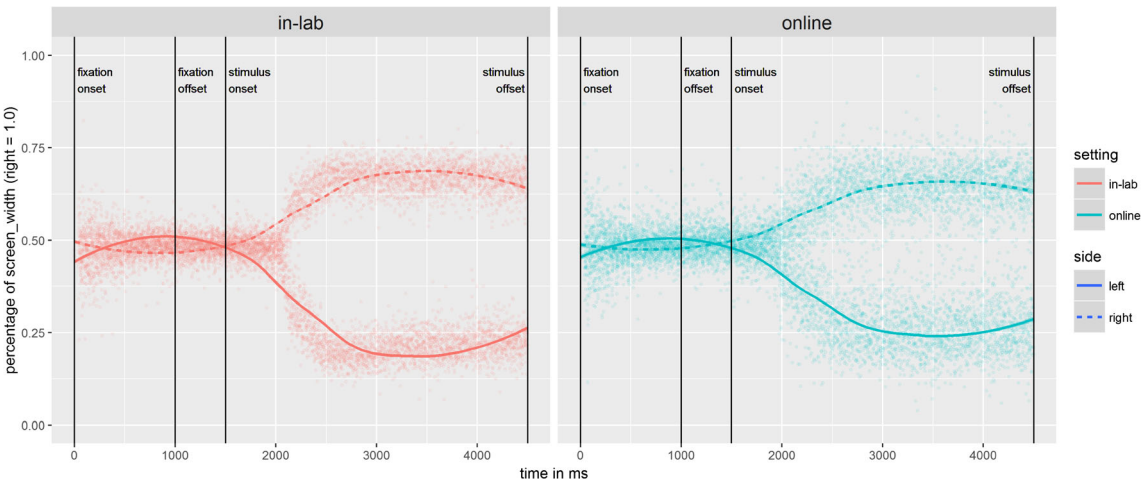


Fig. 8 Free-viewing task side differentiation. Free-viewing task data are represented per experimental setting (in-lab on the left side, online on the right) and line coded into target image screen side (solid on the left side of the screen, dashed being on the right side). Each dot represents a single gaze estimation sample. In both settings a clear preference for the target side can be observed by the change from about the fixation cross (position centrally at 50% of screen size) towards either 25% of screen size (target position left) or 75% of screen size (target position right)

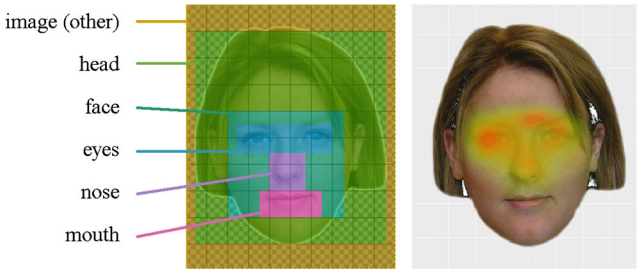


Fig. 9 Regions of interest (ROIs) and heatmap of exemplary face image. The left image shows the ROIs that we defined in 5% steps of the whole image. On the right side, the same face is shown with a heatmap of gaze estimations of on-face ROIs, averaged over all in-lab subjects and both screen sides

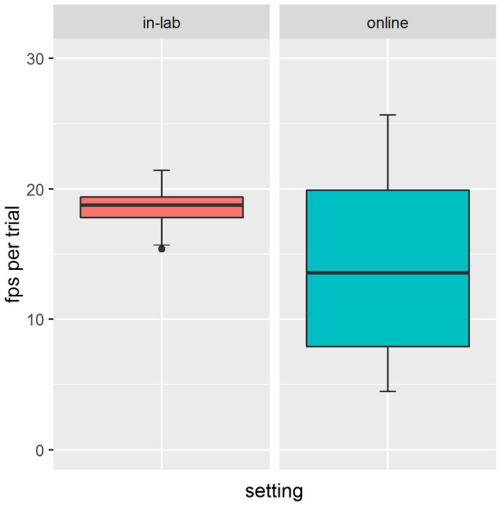


Fig. 11 Frame rate per setting. Boxplots for in-lab (red) and online (cyan) frames per second data with the standard deviations plotted as whiskers

web technology in the acquisition of eye-tracking data online. We employed three paradigms – a fixation task, a pursuit task, and a free-viewing task – and measured each of them in a classical in-lab setting and online (through a crowdscience approach). We found the expected viewing patterns (fixations, saccades, and ROIs) consistently matched our paradigms, with an offset of about 191 px (15% of screen size, 4.38° visual angle) in-lab, whereas online data was found to exhibit a higher variance, lower sampling rate, and longer experimental sessions, but not showing a significant difference in accuracy (offset 211 px, 18% of screen size).

Accuracy of the approach Here we discuss the individual findings from the most basic to the most detailed (and hence across the three tasks that we employed). In the free-viewing task, we were clearly able to determine whether a participant was making a saccade towards the target screen side. This very basic result did not show a significant difference between experimental settings, thus arguing for comparable quality of in-lab and online data. From there, the fixation task results were able to show that we can clearly detect saccades with an average duration of about 450 ms in the in-lab and 750 ms in the online case, reducing the Euclidean distance by 305 px and resulting in an offset of about 188 px (3.94°), whereas commercial systems reach an accuracy of e.g., 0.15° visual angle (EyeLink 1000, e.g., Dalmaijer, 2014).

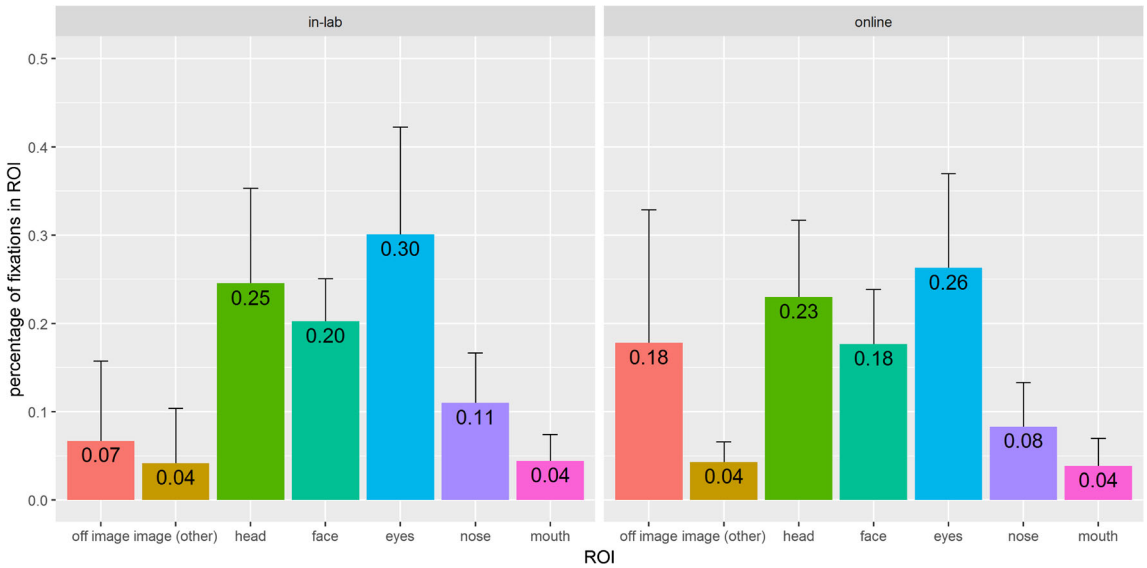


Fig. 10 Free-viewing task results. Each bar denotes the percentage of fixation estimation samples per region of interest (ROI) with the standard deviations plotted as whiskers. In both settings (in-lab, left and online, right), the eyes get predominately fixated compared to other key regions like the nose or the mouth

Together, these two results allow the inferernce that saccades can be identified through online webcam-based eye tracking.

The findings are furthered by the pursuit task. We showed that after the motion started, the spatial offset did not increase but was constant at about 214 px, 4.90°, thus indicating that the participant followed the motion and the pursuit of the target was identified correctly by our application. Again, we did not find a difference in accuracy between experimental settings, putting online data acquisition on par with in-lab data quality. On the other hand, analyzing the speed of eye movements, it becomes apparent that there are still challenges in absolute accuracy when using this methodology. By recording speeds around 37.89°/s, averaged over settings, the measurements indicate a faster movement than the target stimuli (mean 20.41°/s). We see three potential reasons for this deviation. First, due to the fact that we did not smooth our data, microsaccades (e.g., Møller, Laursen, Tygesen, & Sjølie, 2002) could be accountable for these high speeds. Second, due to measurement inaccuracies of the application, the variance of gaze estimations of the same target position creates artificial local eye movements by recording different fixation points, despite the participant keeping his fixation constant. Third, there might be a conflict between presentation and recording on the same computer system, which usually is avoided by using different machines in an in-lab setting. Very carefully designed experiments might be able to pinpoint the exact reason behind the deviation, but our paradigm is not suited for this specific task.

Through the third task we were able to identify specific, attention-guided fixations by introducing a semantic layer in the free-viewing task. Here, we were able to replicate the common finding that Western observers use the eyes of faces as a very distinct feature to analyze (Blais et al., 2008; Caldara et al., 2010; Janik, Wellens, Goldberg, & Dell'Osso, 1978). The significant differences in fixations were as expected: participants pre-dominantly fixated the eye region when compared to other facial landmarks such as the nose or the mouth. Comparing online and in-lab data, we only found an increase in number of fixations for off-image areas in the online data collection. This is in line with our other results that online data quality is lower and noisier than in-lab conducted data, and participants take more time, but argues for no difference in semantic interpretation of stimuli between settings.

Taken together, our results show that using a JavaScript-based online application combined with consumer-grade webcams one can not only estimate saccades up to a target position accuracy of about 200 px and can follow moving stimuli, but that the approach is sensitive enough to identify which parts of a human face carry attentional value to humans, thereby reproducing prior findings.

Improvements With regard to the level of accuracy of online JavaScript-based eye tracking, one has to keep in mind some facts about this study: First, all analyses were performed on raw, unfiltered, and unsmoothed data. Using common eye-tracking algorithms and outlier removal to improve the data obviously would provide an increase in accuracy, especially in those cases in which the sampling rate is low and/or online participants are sitting in a sub-optimal environment (e.g., darkness or too far away). For further studies, we would recommend limiting participants to specific webcam quality-levels and performance of their home computers to achieve an appropriate sampling rate and hence to increase data quality. This approach is in line with work from Xu et al. (2015), who required participants to have a webcam resolution of at least 1,080 × 720 px and achieve at least 25 fps; some of our participants had sampling rates as low as 5 fps, thereby seriously skewing the resolution.

With regard to the free-viewing task, we were rather strict when defining the ROIs of our images in the free-viewing task. Pelphrey et al. (2002), for example, defined 27% of the image as key feature regions, while in our case we only defined 14%. Thus, increasing the ROIs to more liberal values might compensate for the slight inaccuracies of fixation estimation. Lastly, we assume that technical control mechanisms, like head movement detection or instructions on how to use a makeshift chin rest could improve the data quality even further and should be implemented.

Further studies Next to these obvious improvements, further questions are opened up due to the novelty of this research method. For example, not much is known about how many calibration and validation trials are necessary to achieve a high level of gaze estimation accuracy. In our experiments, we chose the values arbitrarily and therefore the calibration/validation-phase took up to 50% of the experimental time of about 30 min. If the validation phase could be minimized or even intrinsically performed (e.g., a similar method to what Papoutsaki et al. are using), the comfort of participating in such studies would be increased dramatically. Gamification of the experimental paradigms would help this task. Another critical question that will need further investigation is the absolute accuracy of the gaze estimation. We found it to be about 200 px (4.16 °), which is in line with around 215 px from Papoutsaki et al. (2016). A much higher accuracy has been achieved by Xu et al. (1.06–1.32 °), due to their use of a headrest and post-hoc processing to detect fixations (meanshift clustering) of three subjects, instead of relying on raw-gaze estimation data. This approach obviously cancels out systematic noise, like we have shown in fixation-

task analysis, which results in a lower nominal offset (around 60 px, 6% of screen size in our case). Yet, we assume that with very careful calibration, instructions, and pre-selections, these values might be improved. This question could be investigated by combining classical in-lab studies with common eye-tracking hardware and web technology studies to have a very clear differentiation between attentional offset (e.g., fixating the forehead vs. the eyes) and measurement inaccuracy (e.g., classical eye-tracking hardware indicates eye region, but the web technology implementation indicates forehead due to inaccuracy). Another potentially interesting factor is indicated by our findings, which show a lower accuracy in the lower portion of the screen. We assume that this is due to the fact that most webcams are positioned on top of the display, yet a specifically designed experiment about spatial accuracy resolution would be necessary to pinpoint potential effects. To ease the process of data acquisition and improve the comfort for participants, it would also be very interesting to re-integrate the automatic calibration through mouse movement and/or clicks by Papoutsaki et al. (2016). This approach would allow for a much more gamified version of online eye tracking, thereby potentially improving the number of interested participants and reducing the number of dropouts throughout the study.

Conclusion

Summing up, we used a JavaScript-based eye-tracking library and consumer-grade webcams to record the eye movements and gaze points of participants through three tasks. We found that the spatial accuracy is about 200 px (17% of screen size, 4.16° visual angle in the in-lab condition), it is consistent over moving stimuli, and accurate enough to replicate the results on human face fixation relations. Comparing in-lab and online conducted data revealed evidence of higher variance, lower sampling rate, and increased experimental time in the online setting, but no significant difference with regard to spatial accuracy in comparison to the in-lab setting. However, these potential influences might be decreased by several points by improving the accuracy by employing instructional, computational, and design-dependent advancements or most notably limiting participation to high performance hardware. While those advances would also help in-lab acquisition, we think that both settings already allow for useful data acquisition in its current state.

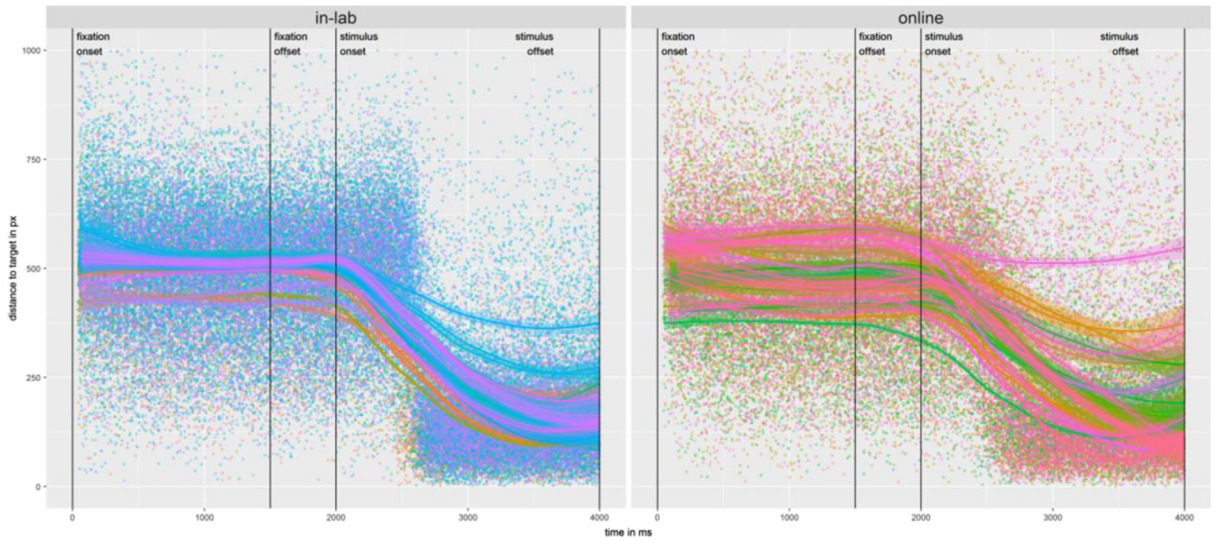
At the current state, we would argue that only those studies that require a very detailed spatial resolution of fixations (e.g., studies in reading, or the dissection of singular items in a crowded display), very time-sensitive information (e.g., high spatio-temporal resolution), or a very short number of trials (e.g., one-trial paradigms) cannot be conducted online. Large ROIs or sections of the screen should be possible to observe. A limiting factor, especially when talking about the temporal course of attention allocation, would be the participant's system performance. We recommend testing the performance beforehand by determining the fps and either end the experiment beforehand or exclude those participants before analyses. Additionally, the resolution of the webcam should be sufficient, but the limits are not yet detected. It is also important to determine which browsers (currently Firefox and Chrome) and what hardware (currently only tested on desktop computers and laptops, but it should be usable on tablets and mobile phones as well) should be supported by the experiment. Cross-browser functionality increases participation rates, but might lead to higher code maintenance. In general, to achieve a higher general data quality, very explicit instructions should be given: No head movement, only eye movement, suitable distance to the screen, and good illumination are just a few, as can be seen at the instructional figure above. Moreover, we would argue for implementing a head-tracking system that restarts calibration once too much movement is detected. Again, it should be decided whether internal (e.g., using mouse movements) or external (e.g., calibration points) calibration is applicable: In most studies, external calibration should allow for a more exact calibration and mediating the seriousness of the experiment, while some studies might need to rely on internal calibration (e.g., external would be too exhausting for the population or to increase gamification aspects).

Taken together, obviously there is a long road ahead of perfectly reliable and accurate online web technology-based eye tracking, yet with these results in a first investigation we do think that it is one worth travelling. The ease of access to participants, rapid data collection, diversity of demographics, and lower cost and time investments are just a few of the factors to consider when deciding on online data collection. Furthermore, even now the spatial resolution should be sufficient for many eye-tracking studies that do not need pixel-wise but rather area-wise accuracy. As the foundation of online experimentation grows, we estimate that algorithms and software will develop in turn and golden standards will emerge that will improve the accuracy towards comparable levels to classical methodological acquisitions. Thus, we think it is worth employing these (still experimental) methods to broaden the possibilities in psychological data acquisition.

Acknowledgements We would like to thank Astrid Hönekopp and Alexander Diel for help in collecting the data and Katharina Sommer for providing the instructional images. All code, raw data and analysis files can be found at The Open Science Framework (https://osf.io/jmz79).

Appendix

Fixation task single subject graph



Averaged distance overview

condition	setting	distance in px	distance in %
dot_20%_20%	in-lab	41.388206	0.04411231
dot_20%_50%	in-lab	-1.299234	-0.01131887
dot_20%_80%	in-lab	-84.998147	-0.11301041
dot_50%_20%	in-lab	9.205665	0.02135174
dot_50%_80%	in-lab	-136.693593	-0.13535110
dot_80%_20%	in-lab	-35.076124	-0.00981185
dot_80%_50%	in-lab	-105.416564	-0.07975572
dot_80%_80%	in-lab	-239.121147	-0.20571218
dot_20%_20%	online	134.324779	0.11492351
dot_20%_50%	online	49.665693	0.01324963
dot_20%_80%	online	-41.268650	-0.10729886
dot_50%_20%	online	36.985062	0.04848134
dot_50%_80%	online	-149.944186	-0.16745254
dot_80%_20%	online	-25.808421	0.01820289
dot_80%_50%	online	-152.182973	-0.11175963
dot_80%_80%	online	-286.191201	-0.25664720

Fixation task descriptive results (Euclidean distance)

condition	setting	distance in %	distance in px
dot_20%_20%	in-lab	0.1541387	184.3716
dot_20%_50%	in-lab	0.1238191	150.2430
dot_20%_80%	in-lab	0.1783645	197.3743
dot_50%_20%	in-lab	0.1294544	144.0125
dot_50%_80%	in-lab	0.1522931	165.1922
dot_80%_20%	in-lab	0.1378035	161.5744
dot_80%_50%	in-lab	0.1267098	156.1566
dot_80%_80%	in-lab	0.1824733	209.7269
dot_20%_20%	online	0.1803735	204.7487
dot_20%_50%	online	0.1670030	199.0196
dot_20%_80%	online	0.2103699	223.7746
dot_50%_20%	online	0.1622174	172.1001
dot_50%_80%	online	0.1898449	190.4822
dot_80%_20%	online	0.1838057	204.1192
dot_80%_50%	online	0.1678164	202.6905
dot_80%_80%	online	0.2292735	256.4242

Fixation task post hoc t-test results

condition 1	condition 2	p value
dot_20%_20%	dot_20%_50%	1.00000
dot_20%_20%	dot_20%_80%	0.74018
dot_20%_20%	dot_50%_20%	1.00000
dot_20%_20%	dot_50%_80%	1.00000
dot_20%_20%	dot_80%_20%	1.00000
dot_20%_20%	dot_80%_50%	1.00000
dot_20%_20%	dot_80%_80%	0.06639
dot_20%_50%	dot_20%_80%	0.01841
dot_20%_50%	dot_50%_20%	1.00000
dot_20%_50%	dot_50%_80%	0.81608
dot_20%_50%	dot_80%_20%	1.00000
dot_20%_50%	dot_80%_50%	1.00000
dot_20%_50%	dot_80%_80%	0.00054
dot_20%_80%	dot_50%_20%	0.01844
dot_20%_80%	dot_50%_80%	1.00000
dot_20%_80%	dot_80%_20%	0.39092
dot_20%_80%	dot_80%_50%	0.02798
dot_20%_80%	dot_80%_80%	1.00000
dot_50%_20%	dot_50%_80%	0.81608
dot_50%_20%	dot_80%_20%	1.00000
dot_50%_20%	dot_80%_50%	1.00000
dot_50%_20%	dot_80%_80%	0.00055
dot_50%_80%	dot_80%_20%	1.00000
dot_50%_80%	dot_80%_50%	1.00000
dot_50%_80%	dot_80%_80%	0.39092
dot_80%_20%	dot_80%_50%	1.00000
dot_80%_20%	dot_80%_80%	0.02731
dot_80%_50%	dot_80%_80%	0.00098

Free-viewing task descriptive values

roi	setting	percentage	sd
eyes	in-lab	0.30066694	0.12169632
eyes	online	0.26283858	0.10678863
face	in-lab	0.20260621	0.04785487
face	online	0.17671009	0.06220009
head	in-lab	0.24523776	0.10776496
head	online	0.22986275	0.08687942
image (other)	in-lab	0.04162069	0.06236636
image (other)	online	0.04299691	0.02282451
mouth	in-lab	0.04433172	0.03005968
mouth	online	0.03866397	0.03136960
nose	in-lab	0.11005767	0.05652128
nose	online	0.08314883	0.05001170
off image	in-lab	0.06699563	0.09069122
off image	online	0.17810328	0.15083662

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(38), 419–439. doi:10.1006/jmla.1997.2558

Birnbaum, M. H. (2000). Introduction to psychological experiments on the internet. In M. H. Birnbaum (Ed.), *Psychological experiments on the internet* (pp. xv–xx). Academic Press. doi:10.1016/B978-012099980-4/50001-0

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3*(8). doi:10.1371/journal.pone.0003022

Boraston, Z., & Blakemore, S.-J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of Physiology, 581,* 893–898. doi:10.1113/jphysiol.2007.133587

Bulling, A., & Gellersen, H. (2010). Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing, 9*(4), 8–12. doi:10.1109/MPRV.2010.86

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods, 42*(1), 286–291. doi:10.3758/BRM.42.1.286

Caldara, R., Zhou, X., & Miellet, S. (2010). Putting culture under the "Spotlight" reveals universal information use for face recognition. *PLoS ONE, 5*(3), 1–12. doi:10.1371/journal.pone.0009708

Chapman, P., Underwood, G., & Roberts, K. (2002). Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F: Traffic Psychology and Behaviour, 5*(2), 157–167. doi:10.1016/S1369-8478(02)00014-1

Chen, M. (2001). What can a mouse cursor tell us more? Correlation of eye/mouse movements on web browsing. *Proceedings of the ACM Conference on Human Factors in Computing Systems,* 281–282. doi:10.1145/634067.634234

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*(35), 12629–12633. doi:10.1073/pnas.0506162102

Dalmaijer, E. S. (2014). Is the low-cost EyeTribe eye tracker any good for research? *PeerJ PrePrints, 4*(606901), 1–35. doi:10.7287/peerj.preprints.141v2

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2013). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, (February 2016), 1–16. doi:10.3758/s13428-013-0422-2

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455–470. doi:10.3758/BF03195475

Fukushima, K., Fukushima, J., Warabi, T., & Barnes, G. R. (2013). Cognitive processes involved in smooth pursuit eye movements: Behavioral evidence, neural substrate and clinical correlation. *Frontiers in Systems Neuroscience, 7*(March), 4. doi:10.3389/fnsys.2013.00004

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review, 19*(5), 847–857. doi:10.3758/s13423-012-0296-9

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2000). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *The American Psychologist, 59*(2), 93–104. doi:10.1037/0003-066X.59.2.93

Holzman, P. S., Proctor, L. R., & Hughes, D. W. (1973). Eye-tracking patterns in schizophrenia. *Science (New York, N.Y.), 181*(4095), 179–181.

Janik, S. W., Wellens, A. R., Goldberg, M. L., & Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces.

Perceptual and Motor Skills, 47, 857–858. doi:10.2466/pms.1978.47.3.857

Lisberger, S. G., Morris, E. J., & Tychsen, L. (1987). Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annual Review of Neuroscience, 10*(1), 97–129. doi:10.1146/annurev.ne.10.030187.000525

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. doi:10.3758/s13428-011-0168-7

Møller, F., Laursen, M. L., Tygesen, J., & Sjølie, A. K. (2002). Binocular quantification and characterization of microsaccades. *Graefe's Archive for Clinical and Experimental Ophthalmology, 240*(9), 765–770. doi:10.1007/s00417-002-0519-2

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer : Scalable webcam eye tracking using user interactions. *International Joint Conference on Artificial Intelligence*.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 32*(4), 249–261. doi:10.1023/A:1016374617369

Semmelmann, K., & Weigelt, S. (2016). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*. doi:10.3758/s13428-016-0783-4

Stewart, N., Ungemach, C., Harris, A. J. L., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers. *Judgment and Decision making, 10*(5), 479–491. doi:10.1017/CBO9781107415324.004

Valenti, R., Staiano, J., Sebe, N., & Gevers, T. (2009). *Webcam-based visual gaze estimation* (1), pp. 662–671.

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95–99. doi:10.1016/j.learninstruc.2009.02.009

Walker, R., Walker, D. G., Husain, M., & Kennard, C. (2000). Control of voluntary and reflexive saccades. *Experimental Brain Research, 130*(4), 540–544. doi:10.1007/s002219900285

Wedel, M., & Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science, 19*(4), 297–312. doi:10.1287/mksc.19.4.297.11794

Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015). TurkerGaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv Preprint arXiv: ...*, *91*(12), 5. Retrieved from http://arxiv.org/abs/1504.06755

Yarbus, A. L. (1967). Eye movements and vision. *Neuropsychologia, 6*(4), 222. doi:10.1016/0028-3932(68)90012-2