Bringing Sexy (Webcam Eye-tracking) Back into the lab

Jason Geller[1][*] ![ORCID], João Veríssimo[2] ![ORCID] & Julia Droulin[3] ![ORCID]

[1]Boston College, Department of Psychology and Neuroscience
[2]University of Lisbon, School of Arts and Humanities
[3]Univetsity of North Carolina - Chapel Hill, Speech and Hearing Divison

Webcam-based eye-tracking offers a scalable and accessible alternative to traditional lab-based systems. While recent studies demonstrate that webcam eye-tracking can replicate canonical effects across domains such as language, memory, and decision-making, questions remain about its precision and reliability. In particular, spatial accuracy, temporal resolution, and attrition rates are often poorer than those observed with research-grade systems, raising the possibility that environmental and hardware factors introduce substantial noise. The present registered report directly tests this hypothesis by bringing webcam eye-tracking back into a controlled laboratory setting. In Experiment 1, we examine the effect of webcam quality (high vs. low) in a single word Visual World Paradigm (VWP) task, testing whether higher-quality webcams yield stronger competition effects, earlier effect onsets, and reduced attrition. In Experiment 2, we assess the impact of head stabilization (chinrest vs. no chinrest) under identical environmental conditions. Together, these studies isolate the causal influence of hardware and movement on webcam eye-tracking data quality. Results will inform a more methodological understanding of webcam-based eye-tracking, clarifying whether its current limitations are intrinsic to the technology or can be mitigated through improved hardware and experimental control.

Keywords: Webcams, Eye-tracking, VWP, Lab, Competition, Spoken word recognition
Words: 2119

Online experimentation in the behavioral sciences has advanced considerably since its introduction at the Society for Computers in Psychology (SCiP) conference in Chicago, IL, in the mid-1990s (Reips, 2021), and its use in the behavioral sciences has grown substantially in the years since. One methodological domain that has shown particular promise in moving online is eye tracking. Traditionally, eye-tracking studies required controlled laboratory settings equipped with specialized and costly hardware—a process that is both resource- and time-intensive. More recently, however, a growing body of research has shown that eye tracking can be successfully adapted to online environments (Bogdan et al., 2024; Bramlett & Wiener, 2024; Cruyssen et al., 2023; James et al., 2025; Prystauka et al., 2024; Slim et al., 2024; Slim & Hartsuiker, 2023; Vos et al., 2022; Yang & Krajbich, 2021; Özsoy et al., 2023). By leveraging standard webcams, researchers can now record eye movements remotely, making it possible to collect data from virtually any location at any time. This shift not only enhances scalability but also broadens access to more diverse and representative participant samples.

Webcam-based eye tracking is straightforward to implement for research purposes. It typically requires only a standard computing device—such as a laptop, desktop computer, tablet, or smartphone—equipped with a built-in or external webcam. Data collection is conducted through a web browser running specialized software that records and estimates eye movements in real time. The accessibility of webcam eye tracking has been greatly enhanced by its integration into several popular experimental platforms, including Gorilla (Anwyl-Irvine et al., 2019), PsychoPy/ Psycho(Peirce et al., 2019), jsPsych (Leeuw, 2014) , and PCIbex ((Zehr & Schwarz, 2022).

To reliably estimate where users are looking, webcam-based eye tracking typically relies on appearance-based methods, which infer gaze direction directly from visual features of the eye region (e.g., pupil and iris appearance). Recent work has extended these methods using deep learning to learn gaze–appearance mappings directly from data (Kaduk et al., 2023; Saxena et al., 2024). This

---

contrasts with research-grade eye trackers, which use model-based algorithms combining infrared illumination with geometric modeling of the pupil and corneal reflections (Cheng et al., 2024).

The most widely used library for webcam eye tracking is WebGazer.js (Papoutsaki et al., 2016; Patterson et al., 2025). WebGazer.js is an open-source JavaScript library that performs real-time gaze estimation using standard webcams. It leverages computer vision techniques to detect the face and eyes, extract image features, and map these features onto known screen coordinates during a brief calibration procedure. Once trained, gaze locations on the screen are estimated via ridge regression (Papoutsaki et al., 2016).

Although webcam eye-tracking is still relatively new, validation efforts are steadily accumulating and the results are encouraging. Researchers have successfully applied webcam-based methods to domains such as language (Bramlett & Wiener, 2025; Geller et al., 2025; Prystauka et al., 2024), judgment and decision-making (Yang & Krajbich, 2021), and memory (James et al., 2025) Overall, these studies replicate canonical effects and show strong convergence with findings from traditional lab-based eye-tracking systems.

However, there are a few shortcomings to web-based eye-tracking. First, effect sizes are often smaller than in lab-based studies (Bogdan et al., 2024; Degen et al., 2021; Kandel & Snedeker, 2024; Slim et al., 2024; Slim & Hartsuiker, 2022; Van der Cruyssen et al., 2024), which typically requires larger samples to achieve comparable power. Second, relative to research-grade eye-trackers, spatial and temporal precision are poorer: webcam approaches can commonly yield spatial accuracy of roughly 4° (Semmelmann & Weigelt, 2018) and temporal resolution can range from 50 ms-1000 ms (Geller et al., 2025; Semmelmann & Weigelt, 2018; Slim et al., 2024; Slim & Hartsuiker, 2023). These constraints make webcam eye-tracking less suitable for research that requires fine-grained spatial or temporal fidelity—for example, designs with many areas of interest (AOI) distributed across the screen or small AOIs [AOIs; James et al. (2025)], or tasks requiring precise moment-to-moment processing (Slim et al., 2024). Lastly, webcam eye-tracking yields high attrition rates. Looking at a number of webcam eye-tracking studies Patterson et al. (2025) found the attrition rate was on average 13%, with some studies reporting attraition as higher as 75% (Geller et al., 2025).

An open question is whether the limitations of web-based eye-tracking stem from the WebGazer.js algorithm itself or from environmental and hardware constraints—and, importantly, whether future improvements can address these issues. On the algorithm side of things, recent work (James et al., 2025) showed that modifying WebGazer.js insofar that the sampling rate is polled consistently and timestamps are aligned to acquisition (when data are received) rather than at completion (when processing finishes) markedly improves temporal resolution. Implementations of these changes in online experiment platforms (e.g., Gorilla and jsPsych) have brought webcam eye-tracking studies closer to the timing fidelity seen in lab-based eye-tracking. For instance, using the Gorilla platform observed a 50 ms timing difference between lab-based effects and online effects and Geller et al noted a 100 ms timing difference

To our knowledge, no study has directly tested how environmental and hardware constraints impact webcam-based eye-tracking data. Slim & Hartsuiker (2023) provided some evidence suggesting that hardware quality may underlie some of these limitations, reporting a positive correlation between webcam sample rate and calibration accuracy. Similarly, Geller et al. (2025) found that participants who failed calibration more often reported using low-quality built-in webcams and working in suboptimal environments (e.g., natural lighting). Together, these findings suggest that both hardware and environmental factors may contribute to the increased noise commonly observed in online eye-tracking data.

Proposed Research

To address environmental and technical sources of noise in webcam eye-tracking, we plan to bring participants into the lab to complete a Gorilla-hosted webcam task under standardized conditions. We will manipulate two factors across two experiments. Experiment 1 will vary webcam quality (high- vs. low-quality external cameras). Experiment 2 varies head stabilization (with vs. without a chinrest). All sessions will be conducted under standardized conditions: identical ambient lighting, fixed viewing distance, the same display/computer model, and controlled network settings. This

design allows us to isolate the causal effects of hardware and movement on data quality. Our key questions are whether higher-quality webcams and reduced head movement decrease noise, thereby (a) increasing effect sizes, (b) yielding earlier onsets of established effects, and (c) reducing calibration failures/attrition rate. As noted above

To examine these factors, we replicate a paradigm widely used in psycholinguistics—the Visual World Paradigm [VWP; Cooper (1974); Tanenhaus et al. (1995)]. The VWP has been used quite successfully with webcam eye-tracking (Bramlett & Wiener, 2024; 2025; Geller et al., 2025; Prystauka et al., 2024). While there are slight variations in how the paradigm is implemented (Huettig et al., 2011), in the version most relevant to the current study, on each trial four images appear in one of four screen quadrants while a spoken word is played. Participants then select the picture that matched the utterance. Related to the current experiment, item sets are sometimes constructed so that the display contains a target (e.g., CARROT), a cohort competitor (CARRIAGE) , a rhyme competitor (PARROT), and an unrelated distractor (TADPOLE). A setup such as this allows one to examine how competition dynamics (i.e., cohort vs. unrelated items or rhyme vs. unrelated items) unfold over the time course of language processing (Colby & McMurray, 2023). Importantly, competition effects have previously been observed in webcam eye-tracking studies using a similar design (Geller et al., 2025) thus serve as good testing case herein.

Experiment 1: High quality webcam vs. low quality webcam

Slim & Hartsuiker (2023) and Geller et al. (2025) observed a relationship between webcam quality and calibration accuracy. Building on this, Experiment 1 will test how webcam quality influences competition effects in a single-word VWP/ Specifically, we ask whether a higher-quality webcam yields (a) a greater proportion of looks (i.e., stronger detectability of competition), (b) earlier emergence of the effect over time, and (c) lower attrition rates relative to a lower-quality webcam.

Hypotheses

We hypothesize several effects related to competition, onset, and attrition.

Competition Effects

(H1a) Participants will show a competition effect, with more looks directed toward cohort competitors than unrelated distractors. (H1b) Webcam quality (high vs. low) will influence the overall proportion of looks, with higher-quality webcams detecting a greater number of looks. (H1c) There will be an interaction between webcam quality and competition, such that the magnitude of the competition effect will be larger in the high-quality webcam condition than in the low-quality condition.

Onset Effects

(H2a) Looks to cohort competitors will emerge earlier than looks to unrelated distractors. (H2b) The onset of looks will occur earlier in the high-quality webcam condition than in the low-quality condition. (H2c) Consequently, the competition effect will emerge sooner in the high-quality webcam condition compared to the low-quality condition.

Attrition (H3) Attrition rates will be lower in the high-quality webcam condition than in the low-quality webcam condition.

Method

Sampling Goal

We conducted an **a priori power analysis** via Monte Carlo simulation in R. Data from 21 participants, collected online using the Gorilla experimental platform during the development of the webgazeR package and employing the same stimuli and VVWP design, were used to seed the simulations. In these data, we observed a cohort effect of approximately 3%. Using this value as our seed, we collapsed the data across time bins to compute binomial counts per trial and fit a binomial generalized linear mixed model (GLMM) to obtain fixed-effect estimates. We then augmented the dataset by adding a between-subjects factor for webcam quality, with participants evenly assigned to high- and low-quality groups. In the high-quality webcam group, we modeled both a higher overall fixation rate and a larger cohort effect, whereas in the low-quality group the cohort effect

was halved relative to the high-quality group. Simulated datasets were generated under this model, and the planned GLMM—including a condition_num × webcam_group interaction—was refit to each simulated dataset (5000). Power was estimated as the proportion of simulations in which the interaction term exceeded $|z| = 1.96$. Analysis scripts are available at [link to be inserted/blinded for review]. Results indicated that a total of **35 participants per group (N = 70)** would provide approximately **90% power** to detect the hypothesized reduction in the cohort effect and overall fixation rate under low-quality webcam conditions. We will therefore recruit participants until we have 35 in each group (N = 70 total). For the calibration analysis (see below), all participants who enter the study will be included.

Materials

VWP

Items. Stimuli were adapted from Colby & McMurray (2023) . Each set comprised four images: a target, an onset (cohort) competitor, a rhyme competitor, and an unrelated item (e.g., rocket, rocker, pocket, bubble). For the webcam study, we used 30 sets (15 monosyllabic, 15 bisyllabic).

Within each set, only the target and its onset competitor served as auditory targets once each, yielding two trial types: TCRU (target–cohort–rhyme–unrelated) and TCUU (target–cohort–unrelated–unrelated). This resulted in 60 trials total (30 sets × 2 targets per set). A MATLAB script generated a unique randomized list per participant, pseudo-randomizing display positions so that each image type was approximately equally likely to appear in any quadrant across subjects.

All 120 images were from a commercial clipart database that were selected by a small focus group of students and edited to have a cohesive style using a standard lab protocol (McMurray et al., 2010)

Auditory Stimuli

Auditory stimuli were recorded by a female monolingual speaker of English in a sound-attenuated room sampled at 44.1 kHz. Auditory tokens were edited to reduce noise and remove clicks. They were then amplitude normalized to 70 dB SPL. . All .wav files were converted to .mp3 for online data collection.

Experimental Setup and Procedure

All tasks will be completed in a single session lasting approximately 30 minutes. The sequence of tasks will be fixed: (1) informed consent, (2) the spoken-word Visual World Paradigm (VWP) task, and (3) a demographic questionnaire.

The experiment will be programmed and administered in Gorilla (Anwyl-Irvine et al., 2019) and conducted on a 23-inch Dell U2312HM monitor (1920 × 1080 px). Participants will wear Sony MDR-7506 headphones to ensure consistent audio presentation and minimize background noise.

To manipulate recording quality, two webcams will be used:

- High-quality condition: A Logitech Brio webcam will record in 4 K resolution (up to 4096 × 2160 px) with a 90° field of view, providing high-fidelity video suitable for accurate gaze estimation and pupil tracking.
- Standard-quality condition: A Logitech C270 HD webcam will record in 720 p resolution, producing video comparable to that of a typical laptop webcam, thereby simulating lower-quality online recordings.

Both webcams will be mounted in a fixed position above the monitor to maintain consistent framing across participants. Lighting will be standardized to ensure uniform image quality across all sessions.

At the start of the study, participants will receive information about the experiment and provide informed consent. They will then adjust their headphone volume to a comfortable level while sample noise plays.

Before the main task, an instructional video will demonstrate the calibration procedure. Calibration will occur twice—once at the start and again after 30 trials—with up to three attempts allowed each time. In each calibration phase, participants will view nine calibration targets and five validation points, looking directly at each target as instructed. Participants will then complete

four practice trials to familiarize themselves with the task. Each trial begins with a 500 ms central fixation cross, followed by a preview display of four images located in the screen's corners. After 1500 ms, a start button appears at the center; participants click it to confirm fixation before hearing the spoken word. The images remain visible throughout the trial, and participants indicate their response by clicking the image corresponding to the spoken target. A response deadline of 5 seconds will be used. Eye movements are recorded continuously during each trials. Following the main VWP task, participants will complete a brief demographic questionnaire, after which they will be thanked for their participation.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Bogdan, P. C., Dolcos, S., Buetti, S., Lleras, A., & Dolcos, F. (2024). Investigating the suitability of online eye tracking for psychological research: Evidence from comparisons with in-person data using emotion–attention interaction tasks. *Behavior Research Methods*, *56*(3), 2213–2226. https://doi.org/10.3758/s13428-023-02143-z

Bramlett, A. A., & Wiener, S. (2024). The art of wrangling. *Linguistic Approaches to Bilingualism*. https://doi.org/https://doi.org/10.1075/lab.23071.bra

Bramlett, A. A., & Wiener, S. (2025). The Art of Wrangling: Working with Web-Based Visual World Paradigm Eye-Tracking Data in Language Research. *Linguistic Approaches to Bilingualism*, *15*(4), 538–570. https://doi.org/10.1075/lab.23071.bra

Cheng, Y., Wang, H., Bao, Y., & Lu, F. (2024, April). *Appearance-Based Gaze Estimation With Deep Learning: A Review and Benchmark* (Issue arXiv:2104.12668). https://doi.org/10.48550/arXiv.2104.12668

Colby, S. E., & McMurray, B. (2023). Efficiency of spoken word recognition slows across the adult lifespan. *Cognition*, *240*, 105588. https://doi.org/10.1016/j.cognition.2023.105588

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Cruyssen, I. Van der, Ben-Shakhar, G., Pertzov, Y., Guy, N., Cabooter, Q., Gunschera, L. J., & Verschuere, B. (2023). The validation of online webcam-based eye-tracking: The replication of the cascade effect, the novelty preference, and the visual world paradigm. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02221-2

Geller, J., Prystauka, Y., Colby, S. E., & Drouin, J. R. (2025). Language without Borders: A Step-by-Step Guide to Analyzing Webcam Eye-Tracking Data for L2 Research. *Research Methods in Applied Linguistics*, *4*(3), 100226. https://doi.org/10.1016/j.rmal.2025.100226

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

James, A. N., Ryskin, R., Hartshorne, J. K., Backs, H., Bala, N., Barcenas-Meade, L., Bhattarai, S., Charles, T., Copoulos, G., Coss, C., Eisert, A., Furuhashi, E., Ginell, K., Guttman-McCabe, A., Harrison, E. (., Hoban, L., Hwang, W. A., Iannetta, C., Koenig, K. M., … de Leeuw, J. R. (2025). What Paradigms Can Webcam Eye-Tracking Be Used For? Attempted Replications of Five Cognitive Science Experiments. *Collabra: Psychology*, *11*(1), 140755. https://doi.org/10.1525/collabra.140755

Kaduk, T., Goeke, C., Finger, H., & König, P. (2023). Webcam eye tracking close to laboratory standards: Comparing a new webcam-based system and the EyeLink 1000. *Behavior Research Methods*, *56*(5), 5002–5022. https://doi.org/10.3758/s13428-023-02237-8

Leeuw, J. R. de. (2014). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1–39. https://doi.org/10.1016/j.cogpsych.2009.06.003

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016, July). WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *International Joint Conference on Artificial Intelligence*.

Patterson, A. S., Nicklin, C., & Vitta, J. P. (2025). Methodological recommendations for webcam-based eye tracking: A scoping review. *Research Methods in Applied Linguistics*, *4*(3), 100244. https://doi.org/10.1016/j.rmal.2025.100244

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Prystauka, Y., Altmann, G. T. M., & Rothman, J. (2024). Online eye tracking and real-time sentence processing: On opportunities and efficacy for capturing psycholinguistic effects of different magnitudes and diversity. *Behavior Research Methods*, *56*(4), 3504–3522. https://doi.org/10.3758/s13428-023-02176-4

Reips, U.-D. (2021). Web-based research in psychology: A review. *Zeitschrift Für Psychologie*, *229*(4), 198–213. https://doi.org/10.1027/2151-2604/a000475

Saxena, S., Fink, L. K., & Lange, E. B. (2024). Deep Learning Models for Webcam Eye Tracking in Online Experiments. *Behavior Research Methods*, *56*(4), 3487–3503. https://doi.org/10.3758/s13428-023-02190-6

Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451–465. https://doi.org/10.3758/s13428-017-0913-7

Slim, M. S., & Hartsuiker, R. J. (2023). Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods*, *55*(7), 3786–3804. https://doi.org/10.3758/s13428-022-01989-z

Slim, M. S., Kandel, M., Yacovone, A., & Snedeker, J. (2024). Webcams as Windows to the Mind? A Direct Comparison Between In-Lab and Web-Based Eye-Tracking Methods. *Open Mind*, *8*, 1369–1424. https://doi.org/10.1162/opmi_a_00171

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science (New York, N.Y.)*, *268*(5217), 1632–1634. http://www.ncbi.nlm.nih.gov/pubmed/7777863

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the Visual World Paradigm. *Glossa Psycholinguistics*, *1*(1). https://doi.org/10.5070/G6011131

Yang, X., & Krajbich, I. (2021). Webcam-Based Online Eye-Tracking for Behavioral Research. *Judgment and Decision Making*, *16*(6), 1485–1505. https://doi.org/10.1017/S1930297500008512

Zehr, J., & Schwarz, F. (2022). PennController for Internet Based Experiments (IBEX). *Open Science Framework*. https://doi.org/10.17605/OSF.IO/MD832

Özsoy, O., Çiçek, B., Özal, Z., Gagarina, N., & Sekerina, I. A. (2023). Turkish-German heritage speakers' predictive use of case: webcam-based vs. in-lab eye-tracking. *Frontiers in Psychology*, *14*, 1155585. https://doi.org/10.3389/fpsyg.2023.1155585