Comparing infrared and webcam-based eye tracking in the Visual World Paradigm

Myrte Vos, Sergey Minor, Gillian Ramchand (Arctic University of Norway)
Corresponding author's e-mail address: myrte.vos@uit.no

Visual World eye tracking is a temporally fine-grained method of monitoring attention, making it a popular tool in online sentence processing [7,3]. Since the pandemic has precluded lab-based eye tracking with an infrared eye tracker (IET), **we replicated a recent Visual World IET study on the incremental processing of verb aspect in English using webcam eye tracking (WET)**, to compare the two methods and assess whether WET can serve as an affordable and accessible alternative to IET even for questions probing the time-course of language processing.

Background WET has been available through WebGazer.js, an open source javascript library, since 2016 [5], and a proof-of-concept method study was published in 2018 [6]. A handful of replication studies have since been done [8,2, *a.o.*], and in the past year a variety of user-friendly software tools for conducting WET studies have been developed, but there remain many outstanding questions about the efficacy of this approach. The study we replicated [4] used the Visual World Paradigm to examine whether the English Past Progressive ('imperfective') and Simple Past ('perfective') verb forms drew preferential looks towards snapshots of an event in an ongoing or a completed state, respectively. We found a large preference for the Ongoing Event picture in the Progressive condition starting even before verb offset, and (surprisingly) no preference for either picture in the Simple Past condition. We sought to replicate this robust and intriguing result in a WET version of this study.

Experiment In a two-alternative forced choice task, participants first heard a preamble (e.g. "It was a crisp winter morning") that created a past tense narrative context for the trial. They were then presented with two side-by-side pictures of the same event: one an ongoing version of the event, and one a completed version (Fig. 1), and a test sentence describing the event using either the Past Progressive form of the verb, or the Simple Past form.

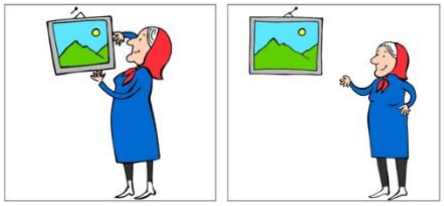
(a)Grandma was hanging a beautiful painting
(b)Grandma hung a beautiful painting

Participants were instructed to choose the picture they thought best matched the sentence. There were 24 test trials and 24 fillers. Filler trials presented pictures of two different events; half the filler sentences included an auxiliary *be* construction describing a completed event. Participants were monolingual English speakers (IET: n=66, recruited in Edinburgh and Trondheim; WET: n=124, recruited on Prolific). The IET experiment was run using an SMI Red 500 eye tracker sampling at 120Hz; the WET experiment was built with jsPsych (v6.3 [1]), mean sampling rate 20.7Hz. Stimuli and counterbalancing were the same in both experiments; in the IET version, participants were calibrated once at the start, while in the WET version, participants re-calibrated once every 12 trials.

Results We successfully replicated both the offline picture choice results (Table 1) and the online gaze pattern results (Fig. 2), to a remarkable degree of similarity. The 'un-perfectivity' of the Simple Past condition was borne out again in the WET version, with preference for the Completed Event picture at 44% in the offline results. In both versions, preference for the Ongoing Event picture in the Progressive condition was at-ceiling, and the aspectual information was processed early enough to be significant before verb offset. We consider this a promising sign that WET can be a suitable method for probing temporally fine-grained processes – one with several practical advantages over IET. Furthermore, this study is (to the best of our knowledge) among the first to be conducted using jsPsych's new (April 2021) eye tracking plugins 'out of box', making the quality of our results easy to replicate using free, open source, beginner-friendly software. Comparative methodological analysis of the two datasets is ongoing, to further our understanding of the factors affecting data quality in WET experiments, but already these results encourage us to treat WET as a serious, exciting alternative to eye tracking in the lab.

	IET		WET	
Event type	**Prog**	**SPast**	**Prog**	**SPast**
Ongoing	95%	46%	98%	56%
Completed	5%	54%	2%	44%

Table 1 Offline responses: percentage of trials in which participants chose a picture type, split by aspectual condition.



(a) Ongoing event (b) Completed

Figure 1 'Grandma hanging a painting'

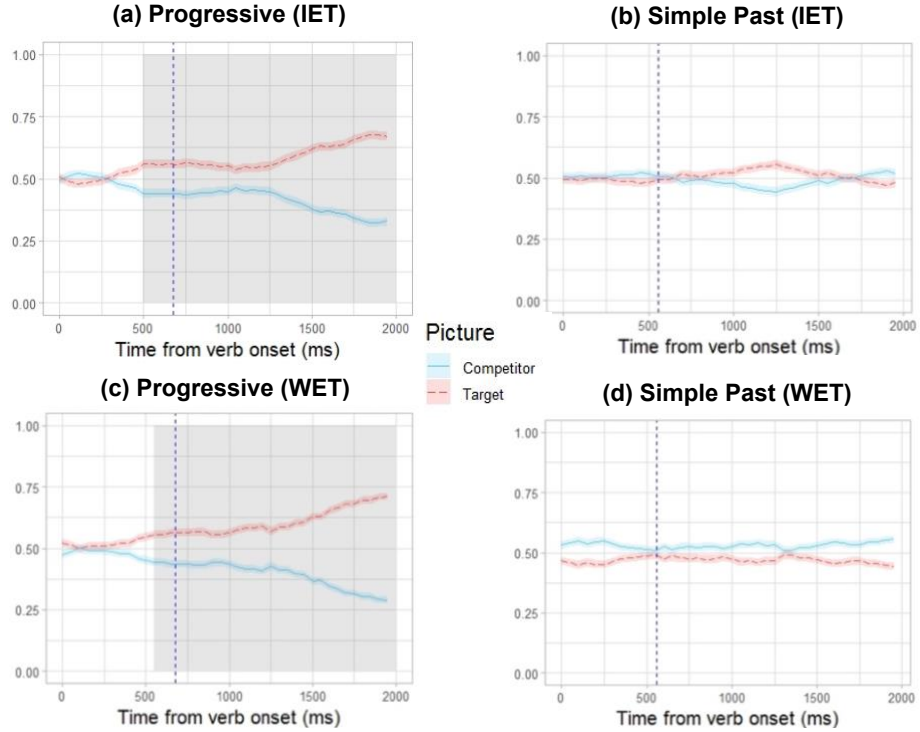


Figure 2 Proportion of looks to the target picture in the Progressive condition (**(a)** and **(c)**, Ongoing Event), and in the Simple Past condition (**(b)** and **(d)**, Completed Event). Data in **(a)** and **(b)** were collected with IET, data in **(c)** and **(d)** with WET. Shading represents the time windows where probability of looks to the target picture was significantly above chance. The dashed vertical line marks average verb offset. Looks to white space were filtered out (IET: 6.36% of gaze data; WET: 27.95%).

Statistical details Offline responses: we fit mixed effects logistic regression models (with random intercepts for subject and item) predicting log-odds of selecting the Ongoing Event picture in the Past Progressive condition (IET & WET: $p < 0.001$), and the Completed Event picture in the Simple Past condition (IET: $p = 0.39$, WET: $p = 0.16$) respectively. Gaze data: trials with above 50% track loss in the IET version, and above 50% gaze predictions outside the screen boundaries in the WET version, were excluded. We used cluster-based permutation analysis (threshold = 0.08) to identify clusters of 50ms time bins where the difference between aspectual conditions was significant. IET Progressive: one cluster from 500 to 2000 ms, sum $z = 103.57$, $p < 0.001$; WET Progressive: one cluster from 550 to 2000 ms, sum $z = 133.18$, $p < 0.001$. IET & WET Simple Past: no significant clusters.

References

[1] J. R. De Leeuw. jspsych: A javascript library for creating behavioral experiments in a webbrowser. Behavior research methods, 47(1):1–12, 2015.

[2] Degen, J., Kursat, L., & Leigh, D. D. (2021). Seeing is believing: testing an explicit linking assumption for visual world eye-tracking in psycholinguistics. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 43.

[3] J. R. Falk Huettig and A. Meyer. Using the visual world paradigm to study languageprocessing: A review and critical evaluation.Acta Psychologia, 137:151–171, 2011.

[4] Minor, S., Mitrofanova, N., Guajardo, G. A., Vos, M., Ramchand, G. (2020, in prep.). Aspect Processing across Languages: A Visual World Eye-Tracking Study.

[5] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays. Webgazer: Scalable webcam eye tracking using user interactions. In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI), pages 3839–3845. AAAI, 2016.

[6] K. Semmelmann and S. Weigelt. Online webcam-based eye tracking in cognitive science: A first look. Behavior Research Methods, 50(2):451–465, 2018.

[7] M. K. Tanenhaus, M. J. Soivey-Knowlton, K. M. Eberhard, and J. C. Sedivy. Integration of visual and linguistic information in spoken language comprehension. Science, 268 (5217):1632–1634, 1995.

[8] X. Yang and I. Krajbich. Webcam-based online eye-tracking for behavioral research. Preprint on PsyArXiv, 2020.