One man's trash is another man's treasure: The need of open data and tools for research reproducibility in screen-based eye-tracking marketing experiments

Vasileios Mygdalis and Nathalie Dens

University of Antwerp, Department of Marketing, Antwerp, Belgium

*Vasileios Mygdalis (Vasileios.Mygdalis@uantwerpen.be) is the corresponding author

One man’s trash is another man’s treasure: The need of open data and tools for research reproducibility in screen-based eye-tracking marketing experiments

Despite the existence of a wide range of eye-tracking experiments in marketing studies, deriving cross-study consensus remains a challenging task. In this article, we identify and discuss some common causes of inter-study variance and propose a set of tools for minimizing cross-study inconsistencies. More precisely, we discuss expected cross-study variances related to how eye-tracking equipment is used, while we also identify other sources of cross-study variance related to picking different eye-tracking variables, differences in the procedure of defining AOI, differences in the experimental protocol and data processing. Based on this review, and inspired by the best practices that have been followed in the computer vision discipline leading to substantial advancements in the past few years, we propose a set of practical and methodological tools that could be taken into consideration for eye-tracking data quality reassurance, maximizing individual study reproducibility and assist in reaching cross-study consensus.

Keywords: Eye-tracking, Attention-based Marketing, Screen-based eye-tracking, tools for reproducibility

Track: Consumer behaviour

1. Introduction

Eye-tracking technology is one of the most important tools for extracting objective metrics to quantify consumer attention in marketing experiments (Orquin & Wedel, 2020; Wedel & Pieters, 2008). The application of eye-tracking technology has assisted in the evaluation of different types of visual stimuli, including different ad designs and product placement options (Ronft et al., 2023), namely banner or native advertising (De Keyzer et al., 2023), personalized advertising (Pfiffelmann et al., 2020), social media advertising (Kohout et al., 2023), (Boerman & Müller, 2022) to name a few. Despite the existence of a wide range of eye-tracking experiments and studies, deriving cross-study consensus still remains a challenging task, since in many cases, data and findings produced by some experimental study are not easy to reproduce in another study, while in addition, it is also not easy to re-use the collected/reported data and/or the derived variables and metrics in meta-analysis or review studies.

In many cases, the employed eye-tracking sensor is assumed as one of the main sources of variance, thus it is almost always reported and detailed in the experimental design part of most eye-tracking papers. Indeed, there exist several types of eye-tracking sensors that have been designed for various experimental settings (indoor/outdoor, screen-based, mobile-based), however, they all feature comparable performance standards in terms of accuracy, precision and sampling rate. The most common experimental variation is indoor screen-based experiments using video-based oculography, which is the main focus of this paper. From the technical standpoint, screen-based experiments offer the best use-case scenario for this type of equipment, since this is when they operate at their highest amount of precision and accuracy, in such a degree that the brand or type of sensor used becomes almost irrelevant, at least from a consumer behaviour analysis standpoint.

From our perspective, we argue that even in an ideal world, assuming the perfect eye-tracking sensor (i.e., zero angle error, no lost signals, infinite sampling rate, etc.), the interpretation of the eye-tracking signal remains the main source of variance between studies. Converting the spatiotemporal eye gaze data into consumer attention variables is not at all a straightforward process; it involves multiple (pre-)/postprocessing steps that define which are the obtained variables and how they are subsequently analysed. The experimental pipeline consists of defining a research question, selecting/designing the visual stimuli material, prototyping the data collection process, appropriate definition of visual Areas of Interest (AOI), extraction of the appropriate eye-tracking variables with respect to the visual stimuli and finally, their analysis.

Besides unavoidable sources of cross-study variance (e.g., different equipment, different research question and different participants in each experiment), there are also exist many "avoidable" sources of cross-study differences. Many of them can be attributed to the lack of a common consensus among studies, i.e., there are too many variations on the names of the standard variables, there are differences on how those standard eye-tracking variables are extracted and calculated, there is a limited amount of raw data sharing, limited use and reporting of principal eye-tracking variables (e.g., fixation time/counts/paths), limited amount of sharing study design material such as visual stimuli and finally, limited amount of analysing and limited reporting of the exact same variables.

In this paper, we identify and discuss some common causes of inter-study variance and propose a set of tools for minimizing cross-study inconsistencies. More precisely, we discuss expected cross-study variances related to how eye-tracking equipment is used, while we also identify other sources of cross-study variance related to picking different eye-tracking variables, differences in the procedure of defining AOI, differences in the experimental protocol and data processing. Based on this review, and inspired by the best practises that have been followed in the computer vision discipline leading to substantial advancements in the past few years, we propose practical and methodological tools that could be taken into consideration for eye-tracking data quality reassurance, maximizing individual study reproducibility and assist in reaching cross-study consensus.

2. Sources of inter-study variance in eye-tracking experiments

In typical screen-based, eye-tracking consumer behaviour studies, recruited participants are asked to view some delivered image/video material on a computer screen while allowing their eye movement to be recorded with some eye-tracking equipment. The objective of the experiment is to capture participant eye movement (i.e., fixations, saccades) with the intention of quantifying consumer attention induced by bottom-up attention stimuli e.g., eye-catching colour/shape patterns, and cognitive top-down attention stimuli, related to voluntary consumer engagement with the ad message/story. The journey to quantify and interpret these aspects involves multiple stages, starting from the capturing the raw eye-tracking sensor data with respect to the provided stimuli.

2.1. Variance introduced by the eye-tracking recording protocol

Modern eye-tracking sensors compute a multi-dimensional signal consisting of 4 core geometric variables that span over time (Georgiadis et al., 2023), namely the estimated relative eye position (X,Y coordinates of eye position on the screen normalized to [0,1]) and pupil size in mm, for both the left and right eye, captured every specific time intervals (i.e., according to the sensor sampling frequency e.g., 120Hz). Assuming the exact same eye-tracking sensor is employed in between experiments, there are still quite a few additional sources of variance that should be taken into account. First, the positional signal needs to be extrapolated from one specific pixel to a wider area in the screen, according to the sensor *spatial resolution* and therefore the important variables that should always be reported are related to scene geometry, e.g., the *distance of the subject/participant to the screen*, the *screen size* and its *resolution*. Second, *sensor calibration* per subject can have a severe impact on the sensor accuracy (Krafka et al., 2016) and appropriate calibration per subject is not always trivial, especially when dealing with non-adults (Zeng et al., 2023). Therefore, besides top-down and bottom-up consumer attention signals, a researcher should always expect a considerable amount of noise, as well, parts of it related to the sensor's nominal accuracy but also heavily moderated by the experimental protocol. In every case, the raw format representation of the signal is of little to no real use to consumer behaviour studies, so it needs to be processed.

2.2 Variance introduced by the definition of the visual stimuli (Areas of Interest)

The raw eye-tracking signal is post-processed in conjunction with the visual content (e.g., the ad) to make it useful in eye-tracking studies, in order to extract the core attention variables e.g., *fixation time, fixation counts, saccade count, dwell time, scanpaths, attention maps* and their variations/aggregations. To this end, researchers are required to define what are the actual visual stimuli on the screen, i.e., the specific *Areas Of Interest* (AOI). In some cases, the provided material as a whole is treated the as the AOI, while in most cases, the visual stimuli is only part of the screen and needs to be annotated. This requires a manual annotation process that should have been straightforward; draw areas that surround the visual stimuli. However, this process is only simple when the visual stimuli is rectangular and the provided material is limited to a few images. When the provided stimuli material consists of many images or even videos, or the visual stimuli are not rectangular, the required annotation effort increases significantly, being both time consuming and prone to annotation error in the process.

No human researcher can be 100% precise at the definition of the AOIs, not due to lack of care in the annotation process, but sometimes because they are also advised to define a little wider areas to compensate for sensor inaccuracies (Orquin et al., 2016). According to technical studies (Vehlen et al., 2022), the definition of wider AOIs should be done with extreme care, as increasing the size might increase fixation recall (number of correctly identified fixations to the AOI) but with the cost of decreasing precision (misclassified non-fixations as fixations). Nevertheless, there is no integrated automated trivial solution for calculating what the exact AOI size should be depending on the sensor, even for rectangular AOIs.

2.3 Variance introduced by the analysed attention variables

Besides variance introduced by the equipment or the AOI definition, which is more or less expected, the computed eye-tracking or consumer attention variables/metrics after raw sensor data processing are still not always the same across studies. Even widely used variables are still subject to vague definitions e.g., one “fixation” might correspond to a temporal sequence of gaze data that span on a specific AOI over a duration of at least 150ms (Ronft et al., 2023), 90ms (Liang et al., 2021), 58ms (van der Laan et al., 2015), or even 300ms (Gheorghe et al., 2023). Studies use various aggregations of the fixations e.g., *average fixation time per subject, total fixation time, time of first fixation*, so it should be expected that aggregating 58ms fixations will produce different results from aggregating 300 ms fixations.

Depending on the research question, different consumer attention variables (or principal variable aggregations) might be more appropriate than others, e.g., time to first fixation is more appropriate for evaluating bottom-up stimuli, while dwell time and attention maps are more appropriate for evaluating top-down ones (van der Laan et al., 2015). In principle, it is always nice to propose new additional variables in the process in order to align with the research question. Nevertheless, if every marketing study proposes their own variations of the principal variables, this hinders conducting cross-study analysis, meta-analysis and does not help in deducting cross study conclusions.

3. Proposals for data analysis prototyping

The expected sources of variance that are mentioned in the previous Section are known in the research community and are in many cases well documented in research articles, stated very clearly in the limitations of the study, or in the description of the experimental protocol. In addition, significant effort has been devoted towards addressing the ethical dimension and there have been many important steps on increasing the availability of data sources. Nevertheless, speaking of open research related to eye-tracking experiments specialized in the marketing domain, we are still at a nascent stage. The main motivation of this article is to examine what additional steps/tools could be considered in the experimental protocol to assist in cross study variance minimization and research reproducibility. The importance of focusing on research reproducibility and open research formats is showcased by other disciplines such as computer vision, which have been enjoying the benefits of open research for quite some time now, sharing research tools, source codes promoting reproducible research and experiencing significant growth in the past few years. Hereafter, we cluster some open research considerations starting from the data collection process, continuing with the data analysis and the reporting process.

3.1 Checks for data acquisition quality assurance

To the best of our knowledge, in most eye-tracking related research articles of the marketing domain, data acquisition typically starts from square one; there is no available existing input data to re-use or compare the present study with. This is due to the fact that the research question in every research article is always different, while researchers mostly focus on comparing or complementing their findings and conclusions with the ones of other research articles, rather than comparing or complementing their input data. However, there exist some useful comparisons with other research articles that do have some things in common that can or should be compared between each other.

Starting from the eye-tracking data, a research article (Zhang & Liu, 2017) that used eye-tracking to assess image quality performed the following input data quality checks: a) Inter-observer-agreement (IOA), i.e., how similar is the variance between the attention maps of different groups of eye-tracking participants, b) saturation, what is the number of participants in each group that adds new information to the aggregated participant group (by analysing the aggregated saliency map), and c) cross-database similarity, i.e., how similar are the attention maps obtained in one study with the ones of another study. Such data quality checks are relevant in almost all eye-tracking experiments and they offer an easy way to reassure that the eye-tracking acquired data between different eye-tracking studies are consistent and trustworthy.

3.2 Data sharing

The best way to promote reproducible research is the “share everything” approach, i.e., create a repository that includes all the obtained raw eye-tracking data, the visual stimuli used in the experiments, the details of the experimental protocol, the software/source code used for the analysis, and perhaps a repository link with instructions on how to obtain and reproduce the results of the article. However, in many cases, researchers rely their analysis on proprietary software that accompanies the employed eye-tracking sensor, thus this practise is perhaps impossible. The main advantage of using proprietary software is that the data collection and analysis processes are streamlined and optimized. Indeed, marketing specialists should not be focused on developing eye-tracking technology/software, but on how to design appropriate research questions, visual stimuli and how to appropriately measure eye-tracking extracted variables. Nevertheless, they should remember that other researchers might not have access to the same equipment/software used in their experiments, so remembering to export the raw obtained gaze data in open data formats is very important (a practice typically allowed and supported by eye-tracking proprietary software providers). Assuming that the visual stimuli and the raw eye-gaze data are available, the remainder processing steps can be reproduced by other researchers even without using the source code or the same analysis tools.

3.3 Open data processing and analysis tools

All the components needed for conducting an eye-tracking experiment exist in free and open-source formats, but only at a scattered fashion. For instance, open source eye-tracking dataset and software (Krafka et al., 2016) (Baltrusaitis et al., 2018) based on webcams and mobile selfie cams is available and is continuously improving (Lai et al., 2023), challenging the accuracy of specialized hardware equipment. In some cases, such solutions are used instead of proprietary hardware/software in recent research articles (Schröter et al., 2021). Open-data practices, regarding at least some parts of the experiment such as sharing the visual stimuli and experimental protocol (Pfiffelmann et al., 2020), (Kohout et al., 2023) are becoming the norm in widely respected academic journals of the field. There are works that share eye-tracking marketing datasets (Liang et al., 2021), or even multimodal datasets that are accompanied with the analysis software (Georgiadis et al., 2023). Open source eye-tracking processing software exists (Dalmaijer et al., 2014) and is maintained in other relevant disciplines. Despite all the above efforts, there is still no marketing specialized streamlined free and open-source solution to dominate the field.

3.4 Overall reproducibility considerations

The raw data acquired during a marketing experiment could be re-used in many studies if they were shared, notably for replicating and complementing existing experiments with additional participants, e.g., for conducting cross analysis in wider regions or meta-analyses. Moreover, when the raw data is shared, this allows for the replication of the study using different definitions of the core variables e.g., fixation of different durations or AOIs of increased/decreased size, or even using a completely different analysis pipeline that will be the norm in the future. We urge the researchers to prioritize raw data sharing over findings/conclusion comparisons.

For the cases where raw data sharing is not possible, due to lack of ethical approval or retraction of the consent of participants, the researchers are advised to report in the Appendices as many possible “core attention” variables as possible, e.g., fixation times, fixations counts, time to first fixation, dwell times, attention heatmaps and scanpaths, even if they are not related to their own research question. For instance, attention heatmaps can be used in other disciplines as well, notably to create and evaluate deep learning methods for saliency detection, irrespective of the marketing research questions addressed.

Finally, whenever a researcher generates their own code or employs open-source repositories for conducting parts of the data analysis, e.g., for statistical tests, it should be advertised and shared with the rest of the community. To conclude, adhering to open-research principles might involve some few additional steps in the overall experimental pipeline process, however, it is typically well rewarding both in article impact (e.g., citation count) and contributes to the worth of the attention-based marketing domain as a whole.

4. Conclusions

This article focused on identifying common causes of inter-study variance other than eye-tracking equipment, such as inconsistencies in eye-tracking variable definition, the importance of precise AOI definition and the detailed description of the experimental protocol. We highlighted the importance of open-data research practices for the potential growth of the domain of attention-based marketing eye-tracking and overviewed some open tools that can be used for re-assuring input data quality between studies and for data analysis.

The main limitation of this study is that the findings are mainly empirical; a proper quantification of the problems mentioned in the previous subsections is lacking, namely what percentage of the eye-tracking research suffers from improper experimental protocol definition, or what are the exact variations of the fixation variables reported in different studies. Secondly, a complete set of tools for streamlining an eye-tracking research study for marketing in a similar manner as proprietary software does can not only exist, but it does exist if we combine and reference all the appropriate repositories. Our future work involves addressing the above limitations and building a repository for connecting all the necessary dots in order to create a competitive open source alternative for conducting and streamlining reproducible eye-tracking experiments.

References

Baltrusaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition* (FG 2018), 59–66.

Boerman, S. C., & Müller, C. M. (2022). Understanding which cues people use to identify influencer marketing on Instagram: An eye tracking study and experiment. *International Journal of Advertising*, 41(1), 6–29.

Dalmaijer, E. S., Mathôt, S., & Van der Stigchel, S. (2014). PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior Research Methods*, 46(4), 913–921.

De Keyzer, F., Dens, N., & De Pelsmacker, P. (2023). The processing of native advertising compared to banner advertising: An eye-tracking experiment. *Electronic Commerce Research*, 23(3), Article 3.

Georgiadis, K., Kalaganis, F. P., Riskos, K., Matta, E., Oikonomou, V. P., Yfantidou, I., Chantziaras, D., Pantouvakis, K., Nikolopoulos, S., Laskaris, N. A., & Kompatsiaris, I. (2023). NeuMa—The absolute Neuromarketing dataset en route to an holistic understanding of consumer behaviour. *Scientific Data*, 10(1), Article 1.

Gheorghe, C.-M., Purcărea, V. L., & Gheorghe, I.-R. (2023). Using eye-tracking technology in Neuromarketing. *Romanian Journal of Ophthalmology*, 67(1), 2–6.

Kohout, S., Kruikemeier, S., & Bakker, B. N. (2023). May I have your Attention, please? An eye tracking study on emotional social media comments. *Computers in Human Behavior*, 139, 107495.

Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., & Torralba, A. (2016). Eye Tracking for Everyone. *2016 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2176–2184.

Lai, B., Liu, M., Ryan, F., & Rehg, J. M. (2023). In the Eye of Transformer: Global–Local Correlation for Egocentric Gaze Estimation and Beyond. *International Journal of Computer Vision*.

Liang, S., Liu, R., & Qian, J. (2021). Fixation prediction for advertising images: Dataset and benchmark. *Journal of Visual Communication and Image Representation*, 81, 103356.

Orquin, J. L., Ashby, N. J. S., & Clarke, A. D. F. (2016). Areas of Interest as a Signal Detection Problem in Behavioral Eye-Tracking Research. *Journal of Behavioral Decision Making*, 29(2–3), 103–115.

Orquin, J. L., & Wedel, M. (2020). Contributions to attention based marketing: Foundations, insights, and challenges. *Journal of Business Research*, 111, 85–90.

Pfiffelmann, J., Dens, N., & Soulez, S. (2020). Personalized advertisements with integration of names and photographs: An eye-tracking experiment. *Journal of Business Research*, 111, 196–207.

Ronft, S., Friedrich, M. G., & Sofiullah, M. (2023). Effect of product placement labeling on visual product and brand reception – an empirical eye tracking study. *Journal of Marketing Communications*, 0(0), 1–16.

Schröter, I., Grillo, N. R., Limpak, M. K., Mestiri, B., Osthold, B., Sebti, F., & Mergenthaler, M. (2021). Webcam Eye Tracking for Monitoring Visual Attention in Hypothetical Online Shopping Tasks. *Applied Sciences*, 11(19), Article 19.

van der Laan, L. N., Hooge, I. T. C., de Ridder, D. T. D., Viergever, M. A., & Smeets, P. A. M. (2015). Do you like what you see? The role of first fixation and total fixation duration in consumer choice. *Food Quality and Preference*, 39, 46–55.

Vehlen, A., Standard, W., & Domes, G. (2022). How to choose the size of facial areas of interest in interactive eye tracking. *PLOS ONE*, 17(2), e0263594.

Wedel, M., & Pieters, R. (2008). A Review of Eye-Tracking Research in Marketing. In N. K. Malhotra (Ed.), Review of Marketing Research (Vol. 4, pp. 123–147). Emerald Group Publishing Limited.

Zeng, G., Simpson, E. A., & Paukner, A. (2023). Maximizing valid eye-tracking data in human and macaque infants by optimizing calibration and adjusting areas of interest. Behavior Research Methods.

Zhang, W., & Liu, H. (2017). Toward a Reliable Collection of Eye-Tracking Data for Image Quality Research: Challenges, Solutions, and Applications. IEEE Transactions on Image Processing, 26(5), 2424–2437.