American Economic Review: Papers & Proceedings 2016, 106(5): 309–313
http://dx.doi.org/10.1257/aer.p20161009

The Power of Eye Tracking in Economics Experiments†

By Joanna N. Lahey and Douglas Oxley*

Eye tracking is a technology that tracks eye activity. Our focus in this paper is on its ability to track where eyes are looking and for how long, which can then be used to inform economic theories of search and screening behavior, including discrimination. Eye tracking can also measure biometric functions related to emotions (such as pupil dilation, Wang, Spezio, and Camerer 2010) and mental effort (using rapid eye movements called saccades), but these functions are outside the scope of this paper.

Surprisingly, eye tracking was first employed for research and marketing purposes in the late nineteenth century using invasive techniques. Continuing technology improvements in the twentieth century led to greater use in psychology and marketing research. The past ten years have seen further improvements in accuracy, ease of use, unobtrusiveness, and, importantly, cost. Low end systems can be purchased at the time of publication for under $5,000, and even state-of-the-art portable trackers with accompanying software can be had for under $30,000. Researchers interested in purchasing eye-tracking equipment should check prices directly as they are continually dropping. These improvements have led to expanded use in many fields. Several literature reviews summarize these research agendas in fields including psychology (e.g., Feng 2011); education and training (Rosch and Vogel-Walcutt 2013); user interface (Bergstrom and Schall 2014); software engineering (Sharafi, Soh, and Guéhéneuc 2015); and even surgery (Hermens, Flin, and Ahmed 2013).

Although eye tracking has been used extensively in marketing (e.g., Wedel 2013), economics has not yet taken full advantage of this technology. A small number of experimental economics papers have used eye tracking (Knoepfle, Wang, and Camerer 2009; Wang, Spezio, and Camerer 2010; Reutskaja et al. 2011). However, even these papers use only a portion of its functionality, focusing primarily on scan-paths (the paths the eyes take) to study learning and consumer choice, or the biometric aspects of eye tracking including pupil-dilation to study truth telling and deception in sender-receiver games.

This paper discusses eye-tracking technology in general and specifically addresses additional areas of economics research that could benefit from the use of eye trackers: labor and personnel economics. We provide general guidelines for researchers interested in utilizing eye-tracking technology in their own work and provide more specific tips using what we have learned from conducting eye-tracking experiments to explore hiring discrimination in resumes.

In our first study using eye-tracking with a professional student sample to study age and race discrimination, we were motivated to find out how discrimination occurs during audit studies (e.g. Bertrand and Mullainathan 2004; Lahey 2008). Eye tracking is able to tell us whether screeners look at resumes, how long they look at resumes, what parts of resumes they look at, and in what order they look at each part. We can then answer questions about whether or not they see a name or an age and then discard the resume, and, if not, what job seekers can do to make their resumes more attractive. Finally, we can use this information to differentiate between theories of discrimination, for example, do participants look longer at black resumes for items that provide information about quality (as with levels-based statistical discrimination) or do they spend less time (as with variance-based statistical discrimination).

* Lahey: Texas A&M, TAMU MS 4220, College Station, TX 77843, and NBER (e-mail: jlahey@tamu.edu); Oxley: Texas A&M, 4102 Amber Trace Court, Sugar Land, TX 77479 (e-mail: dougoxley@gmail.com). Thanks to Matt Notowidigdo for extremely helpful suggestions and to the Sloan foundation Grant #B2012-23 for funding.

† Go to http://dx.doi.org/10.1257/aer.p20161009 to visit the article page for additional materials and author disclosure statement(s).

I. How To

A. *Designing the Experiment*

As with any study, you should start with theory. What do you want to test? Who is your ideal subject pool? What is your ideal setting? Although it is likely that you will have to make compromises, answers to these questions will guide your operationalization choices.

A key operational choice is what type of eye tracker to use. There are three main types of eye-tracking systems: head-mounted systems that participants wear as goggles, stationary desk mounted systems, and small portable systems. Head-mounted systems are useful in any application that requires participants to view their surroundings and are popular in advertising research. The other two types of systems are used in any application that allow the participant to remain stationary. For the student sample of our resume study, we used a stationary desk-mounted system from Applied Science Laboratories (ASL) that sat at the base of a monitor and required a proctor to calibrate the system. However, having a portable system that could travel made it much easier to recruit human resource (HR) managers for our second study exploring age and race discrimination with a population that regularly makes hiring decisions, so we purchased a portable self-calibrating eye tracker from Tobii. Both of these systems included a small black box that sits at the base of the monitor.

In addition to hardware, you will need to purchase software both to run the experiment and to process the data afterward. While most manufacturers will sell the hardware and software in a package, it is possible that you may need additional software. For example, the software that came with the ASL hardware was not sensitive enough to also perform an Implicit Association Test (IAT), so we purchased Inquisit by Millisecond Software so that we could run all of our tests without changing software mid-session. Holmqvist et al. (2011) provide detailed discussions of questions to ask manufacturers about the more complicated characteristics of each type of eye tracker and software system.

In general, there are not many contra-indications when choosing participants. You may want to screen for learning disabilities or native English speaking as they may have different eye-tracking patterns (Holmqvist et al. 2011). Some participants have eyes that are harder to track than others and will thus take more time during the calibration phase. However, items like glasses are generally not a concern.

Freedom from distraction is especially important during eye-tracking experiments. For our student sample, we used a dedicated laboratory. For our field HR sample, we used participant offices or conference rooms and created a quiet booth using a portable clothes rack and blankets at HR fairs. Although good lighting is helpful, perfect lighting is only necessary if very precise measurements are needed, something more common in psycholinguistics research.

If you are in a university setting, it is likely that another researcher on campus has an eye-tracking system that you can borrow to test out or even use for your study. For example, Texas A&M University has several labs already using different types of eye trackers in the psychology, computer science, engineering, and art departments. Indeed, for our first study with the student sample, we used Gerianne Alexander's system housed in her Brain and Gender laboratory along with her expertise and many of her proctors. Working with someone familiar with these systems is invaluable, especially if you can exchange grant money for RA training.

Types of Outputs.—Eye trackers provide a wealth of measurements. The first output of interest does not need eye tracking at all, and is the amount of time spent on each resume screen. The next output of interest is "fixations," which track where the eyes rest for 200 to 300 milliseconds. A scan-path connects fixations in chronological order. In addition, eye trackers can measure pupil dilation and saccades as mentioned above. These measurements can be combined into more complicated measures during analysis (see Holmqvist et al. 2011; Bergstrom and Schall 2014; and Sharafi, Soh, and Guéhéneuc 2015 for more information on outputs).

Factors that Constrain Inputs.—

Precision: As of the time of writing, most affordable eye-tracking equipment is unable to track precisely enough to consistently measure a participant's eye movements while reading standard-size screen fonts that are single spaced

(Spinner, Gass, and Behney 2013). Thus, a study must choose either to track larger regions of text, such as a paragraph, or to use font sizes that may be abnormally large for participants.

A solution to this problem is to define specific "areas of interest" (AOI) or "look zones" that are experimentally meaningful. We broke up our resumes into eight AOI: a box surrounding the name, one surrounding the entire work history except the dates, another surrounding only the dates of the work history, and so on. Current best practices suggest that the boxes be slightly larger than the actual area of interest to allow for imprecision and drift (Sharafi, Soh, and Guéhéneuc 2015). Depending on the type of experiment, you may want these AOI to remain fixed so that participants anticipate where to find specific information.

Calibration: Prior to any measurement, a participant must go through a calibration routine. With our initial lab-based calibration, the proctor guided the participant through the calibration process on the other side of a divider, asking the participant to look at specific areas on the screen. Our portable system self-calibrated with the participants following instructions on the screen to keep their eyes on a bouncing ball. Calibration generally lasts around a minute.

As the study progresses, gradual decline in calibration can occur (Nyström et al. 2013). A few different measures can be taken to reduce the impact of a decline in calibration. First, eye-tracking tasks should be placed in the study shortly following a calibration. Second, in systems that allow it, a study proctor should monitor diagnostic data from the equipment, but in a manner that avoids the Hawthorne effect, especially when social desirability bias may play a role. Third, the study might want to include breaks for recalibration as necessary. We have found that the heuristic of keeping the eye-tracking task to under 30 minutes to be a reasonable one.

B. *Data*

Analysis.—Eye-tracking data collection is not yet perfect. Many of the eye-tracking hardware solutions will have the ability to approximate the reliability of a given data point based on a temporary loss of tracking of the pupil or cornea. This leads to a particular eye-tracking data point being marked into three categories: best, good, and mediocre. The heuristic we used was that 80–100 percent tracking is best, while 60–80 percent is good. This percent tracking statistic can and should be used during proctor training with systems that require proctor calibration to make sure that proctors are applying the protocol correctly.

Most eye-tracking software can combine the raw time, fixations, and scan-path data to more useful outputs for analysis. "First look" is literally the first look that a participant takes. In our resume study, the first look was usually the name at the top of the resume. It may also be useful to collect information on the amount of time spent and the number of fixations in each AOI. We use these data to test theories of discrimination. In addition to showing the scan-path, most software can provide heat maps as another way to visualize the full data. Scan-paths can also be used in conjunction with AOI to create transition matrices that are informative about economic theories. For example, after looking at one AOI, where do they look next? Alternatives to the proprietary software exist, e.g., the Python programming language and the R project, although these require more advanced programming skills.

Management.—Another factor that must be accounted for in an eye-tracking study is the anticipated pipeline of data and how it will be filtered. For many studies, it is expected that 60 or 120 data points will be collected per second. This creates a substantial challenge in data management. If a lab study has 150 participants and ten minutes of recording, a raw data file may be larger than 3,000 megabytes. This file size is not a storage problem for commonly available hardware. It can be a problem, however, for some analytic software to handle files of this size. Similar to the way in which big data analyses occur, we recommend a reduction phase be utilized. This reduction phase aggregates data in meaningful ways prior to connecting the aggregated eye-tracking data to participant level data such as demographics and other participant responses to stimuli. This means that it is critical that the data reduction include participant ID as one of the grouping levels. Other than that, the choice of how to aggregate is primarily dependent upon the research goals.

Both R and Python are capable of aggregating large data files into reduced datasets that would be sized appropriately for deeper statistical analysis. Another technology that can be utilized for aggregation on large datasets is a relational database management system (RDBMS), such as PostgreSQL. The queries and views available in an RDBMS are designed to link and aggregate multiple large datasets in a myriad of ways.

In our initial student study, the standard data analysis software provided by ASL for their eye-tracking equipment did not have the ability to filter/group the data properly for our analysis needs. Therefore, we transferred all of the eye-tracking data and Inquisit data into a Microsoft Access database. In Access, we created several subroutines to acquire data from the large number of Microsoft Excel files that ASL's analysis software output for each participant. In addition, we created several SQL language queries in Access to join and filter the data in useful ways. As the study progressed, it became apparent that Microsoft Access's maximum file size of two gigabytes per individual database file would not be adequate for the raw eye-tracking data. Therefore, we converted the back end storage of data to PostgreSQL with some data management queries residing there and other analysis queries and data still contained in Access. We connected the two databases using ODBC. Finally, once the data had been filtered and joined, we conducted most of our statistical analyses using Stata software and its internal programming language with the exception of some heat map graphs that were generated using Python and MatPlotLib. For our second study using the Tobii tracker, we have been using their off-the-shelf software and Python to port our trimmed datasets to .csv files that are then imported to Stata.

Backups.—As we were utilizing a lab that had multiple studies occurring simultaneously, we were concerned about the potential for data loss. Therefore we instituted a secure backup solution involving routine transfer of the data files off the lab computer and onto a protected network storage server. This proved to be a prescient move as we had a proctor accidentally overwrite a prior participant's data, which we were able to restore from the backup.

II. Going Forward

Eye tracking can also be combined with additional neuroeconomics measures such as skin conductance, cortisol, etc. to get a full emotional profile of the participant. If you wish to broaden your scope to pick up these additional measures, Holmqvist et al. (2011) provides technological considerations for many of these additional types of data collection.

How can you use eye tracking in your research? Eye tracking does not need to be limited to advertisements and resumes. Any procedure that requires people to look or to make decisions can be studied with eye tracking. How people look, how long they take, what paths they take, and so on, can inform economic theory. Eye tracking can open many black boxes. You are only limited by your imagination.

REFERENCES

Bergstrom, Jennifer Romano, and Andrew Schall, eds. 2014. *Eye Tracking in User Experience Design*. Boston: Morgan Kaufmann.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha And Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Feng, Gary. 2011. "Eye Tracking: A Brief Guide for Developmental Researchers." *Journal of Cognition and Development* 12 (1): 1–11.

Hermens, Frouke, Rhona Flin, and Irfan Ahmed. 2013. "Eye Movements in Surgery: A Literature Review." *Journal of Eye Movement Research* 6 (4): 1–11.

Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van De Weijer, eds. 2011. *Eye Tracking: A Comprehensive Guide To Methods and Measures*. Oxford, UK: Oxford University Press.

Knoepfle, Daniel T., Joseph Tao-Yi Wang, and Colin F. Camerer. 2009. "Studying Learning in Games Using Eye-Tracking." *Journal of the European Economic Association* 7 (2-3): 388–98.

Lahey, Joanna N. 2008. "Age, Women, and Hiring: An Experimental Study." *Journal of Human Resources* 43 (1): 30–56.

Nyström, Marcus, Richard Andersson, Kenneth Holmqvist, and Joost Van De Weijer. 2013. "The Influence of Calibration Method and Eye Physiology on Eyetracking Data Quality." *Behavior Research Methods* 45 (1): 272–88.

Reutskaja, Elena, Rosemarie Nagel, Colin F. Camerer, and Antonio Rangel. 2011. "Search Dynamics in Consumer Choice Under Time Pressure: An Eye-Tracking Study." *American Economic Review* 101 (2): 900–926.

Rosch, Jonathan L., and Jennifer J. Vogel-Walcutt. 2013. "A Review Of Eye-Tracking Applications as Tools for Training." *Cognition, Technology, and Work* 15 (3): 313–27.

Sharafi, Zohreh, Zéphyrin Soh, and Yann-Gaël Guéhéneuc. 2015. "A Systematic Literature Review on the Usage of Eye-Tracking in Software Engineering." *Information and Software Technology* 67: 79–107.

Spinner, Patti, Susan M. Gass, and Jennifer Behney. 2013. "Ecological Validity in Eye-Tracking." *Studies in Second Language Acquisition* 35 (02): 389–415.

Wang, Joseph Tao-Yi, Michael Spezio, and Colin Camerer. 2010. "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth-Telling and Deception in Sender-Receiver Games." *American Economic Review* 100 (3): 984–1007.

Wedel, Michel. 2013. "Attention Research in Marketing: A Review of Eye Tracking Studies." Robert H. Smith School Research Paper RHS 2460289.