

Mapping and Correcting the Influence of Gaze Position on Pupil Size Measurements

Taylor R. Hayes and Alexander A. Petrov

Department of Psychology
Ohio State University

Pupil size is correlated with a wide variety of important cognitive variables and is increasingly being used by cognitive scientists. Pupil data can be recorded inexpensively and non-invasively by many commonly used video-based eye tracking cameras. Despite the relative ease of data collection and increasing prevalence of pupil data in the cognitive literature, researchers often underestimate the methodological challenges associated with controlling for confounds that can result in misinterpretation of their data. One serious confound that often is not properly controlled is *pupil foreshortening error (PFE)*—the foreshortening of the pupil image as the eye rotates away from the camera. Here we systematically map PFE using an artificial eye model and then apply a geometric model correction. Three artificial eyes with different fixed pupil sizes were used to systematically measure changes in pupil size as a function of gaze position with a desktop Eyelink 1000 tracker. A grid-based map of pupil measurements was recorded with each artificial eye across 3 experimental layouts of the eye tracking camera and display. Large, systematic deviations in pupil size were observed across all 9 maps. The measured PFE was corrected by a geometric model that expressed the foreshortening of the pupil area as a function of the cosine of the angle between the eye-to-camera axis and the eye-to-stimulus axis. The model reduced the root mean squared error of pupil measurements by 82.5% when the model parameters were pre-set to the physical layout dimensions, and by 97.5% when they were optimized to fit the empirical error surface.

Keywords: Pupillometry, Pupil foreshortening error, Eye tracking, Artificial eye

The human pupillary response correlates with a wide range of important cognitive variables including mental workload (e.g., Hess & Polt, 1964; Kahneman, 1973; Klingner, Tversky, & Hanrahan, 2011), emotional valence (Partala & Surakka, 2003), attention (Beatty, 1982a), working memory (Beatty & Kahneman, 1966), arousal (Murphy, Robertson, Balsters, & O'Connell, 2011), decision-making (Einhäuser, Koch, & Carter, 2010), surprise (Preuschoff, Marius, & Einhäuser, 2011), and uncertainty (Nassar et al., 2012). While the underlying mechanisms that drive cognitive pupillary effects are still an active topic of inquiry, recent data suggests that the pupillary response may reflect noradrenergic activity in the brain (Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Aston-Jones & Cohen, 2005; Koss, 1986; Samuels & Szabadi, 2008). As a result of these recent developments, pupil size is increasingly being used by cognitive scientists as an important measure of cognitive processing.

Pupil data can be collected inexpensively and non-

invasively by most modern image-based eye-tracking systems, but researchers often underestimate the methodological challenges in producing well-controlled studies that allow for cognitive interpretations of the pupillary response (Holmqvist et al., 2011). Most researchers are aware of the most serious confound—the pupillary light reflex. The pupil diameter depends first and foremost on the luminance of the stimulus and the ambient illuminance of the experimental room. Changes in either of these variables can produce strong light reflexes that modulate pupil size by as much as 50% (Miller & Newman, 2005; Loewenfeld, 1993). Cognitive researchers control for this confound by using isoluminant stimuli and maintaining constant room illumination. A lesser known yet equally problematic confound is *pupil foreshortening error (PFE)*. Changes in gaze position produce foreshortening of the pupillary image because the eye tracking camera is fixed but the eye rotates. Specifically, as the eye rotates away from the camera, the pupil image becomes more elliptical and shrinks in apparent area by as much as 10%. This is a large margin compared to the magnitude of the cognitive effects, which rarely exceed 5% change in pupil size. Despite its potential seriousness, the PFE is not corrected in many commonly used remote eye-tracking systems (e.g., Tobii, Tobii Technology AB, 2010, and EyeLink, SR Research, 2010). All too often, the PFE is not controlled by the experimental design either (Brisson et al., 2013). The combination of these factors has led to an increasing number

This article is accepted for publication in *Behavior Research Methods*, April 2015.

Correspondence concerning this article should be addressed to Taylor R. Hayes, Department of Psychology, Ohio State University, Columbus, Ohio 43210. E-mail: taylor.r.hayes@gmail.com

of published studies with inadequate control for PFE.

There are currently two different methods aimed at circumventing PFE using experimental design. The most conservative method—and the one recommended by most commercial eye-tracking providers—is to have participants maintain constant fixation throughout each trial (e.g., SR Research, 2010, p. 98). The constant fixation method is implemented by defining a small fixation boundary area and verifying that the eye position remains within this area throughout the pretrial fixation baseline and stimulus presentation periods. This minimizes PFE by keeping the optical axis of the eye at a fixed angle from the camera. Standard baseline normalization is typically used in conjunction with the constant fixation method, where the task-evoked pupillary response is measured as the absolute difference in millimeters or percent change between the stimulus presentation and pretrial baseline periods (Beatty & Lucero-Wagoner, 2000). However, the constant fixation method has two serious drawbacks. First, it limits the types of tasks that can be studied to those with a constant fixation location. Unfortunately, this rules out studies of important topics such as visual search or reading. Second, it is hard for participants to consistently maintain fixation for long periods of time. In addition, the pupillary response lags behind the stimulus by about 400 ms peaking around 1-2 seconds post-stimulus, which necessitates longer baseline and stimulus durations where participants must maintain fixation (Partala & Surakka, 2003; Murphy et al., 2011; Hayes & Petrov, 2015). This results in a high number of invalid trials that increases as the stimulus duration increases. Another PFE-mitigation method that is sometimes recommended by eye-tracking providers (SR Research, 2010, p. 98) is stimulus-position counterbalancing. In these designs, there are several positions where the stimulus can appear, but the number of presentations is counterbalanced across trials. This method still does not allow the study of tasks that require free viewing of the stimulus. Neither does it allow for comparisons of physical pupil size among different gaze locations because PFE varies across the visual field (Gagl, Hawelka, & Hutzler, 2011). In sum, though effective at mitigating the PFE, both methods have serious limitations.

A completely different approach to the problem is to measure PFE and then correct the pupillometric data prior to analysis (Gagl et al., 2011; Brisson et al., 2013). Gagl and colleagues (2011) recently pioneered this approach with the aid of an artificial eye model with a fixed pupil size. The artificial eye was moved across a single horizontal scan line, emulating a sentence reading task. The PFE could thus be measured explicitly and a correction formula was developed. The horizontal extent of the sentence stimuli spanned from -17° to $+9^\circ$ from the screen center. This resulted in systematic pupillometric error from $+5\%$ to -13% , respectively, relative to the true pupil area of the artificial eye. These data allowed the development of a mathematical model of the PFE. This model was used to correct the empirical data from a sentence reading task and a “Z-string” reading task in which participants scanned words composed entirely of the letter Z. When the measurement error was corrected, previously re-

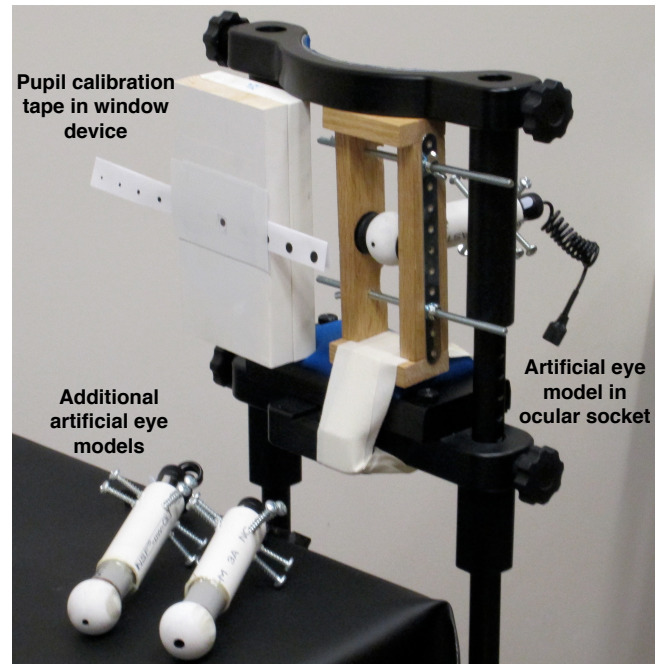


Figure 1. Photograph of the artificial eye, ocular socket, and pupil calibration devices. Each artificial eye was measured in a constant left-eye position within the chin- and forehead-rest. The pupil calibration apparatus was not present during artificial-eye data collection as shown, but rather replaced the ocular socket during the pupil calibration procedure.

ported discrepancies between the pupillary response during word recognition (Kuchinke, Vö, Hofmann, & Jacobs, 2007; Vö et al., 2008) and sentence reading tasks (Just & Carpenter, 1993; Raisig, Welke, Hagendorf, & van der Meer, 2010) were revealed to be artifacts of PFE (Gagl et al., 2011). This pioneering study illustrates the importance of accounting for PFE and how it can lead to incorrect cognitive interpretations if not properly controlled or corrected. One major shortcoming of this study is that it only mapped a single horizontal scan line. The resulting correction formula is applicable to this special case only.

Brisson and colleagues (2013) used an object pursuit task to characterize the PFE of three popular eye-tracking systems (Tobii X120, Tobii T120, and EyeLink 1000) across both horizontal and vertical changes in gaze position. Forty-four human participants tracked a circle moving counterclockwise across the display in an elliptical pattern that covered up to 22° of horizontal visual angle and up to 14° of vertical visual angle (depending on the system). The results showed substantial measurement errors in pupil diameter as a function of gaze position in all three systems. A linear regression model using X and Y gaze coordinates as predictors was able to explain between 9% and 20% of the pupil variance depending on the system and the maximum visual angle. While these data provided important new insights into the extent of PFE with human participants across multiple eye-tracking systems, the regression-based correction procedure only ac-

counted for a small amount of the total PFE. Furthermore, even low-effort tasks such as object tracking require attentional mechanisms and cognitive effort that induce pupillary variability (Alnaes et al., 2014; Beatty, 1982a, 1982b; Kahneman, 1973) that complicates the estimation of the foreshortening error.

In the present article we systematically map and correct pupil foreshortening error across the full display using a set of artificial eyes with fixed, known pupil diameters. We report two studies. The first verified the ratio scale between the “arbitrary units” of the EyeLink 1000 and physical units. The second study systematically maps PFE and develops both a parameter-free and a parameterized geometric model correction that virtually eliminates PFE. Three artificial eyes were measured across the full extent of a 21” monitor in each of three separate experimental layouts varying the relative distances between the camera, monitor, and eye. The measurements revealed large distortions in recorded pupil diameter as a function of gaze position. A simple geometric model that estimates pupil foreshortening as the cosine of the angle between the eye-to-camera axis and the eye-to-stimulus axis produced excellent fits to the data (cf. Atchison & Smith, 2000; Spring & Stiles, 1948). The model reduced the root mean squared error of pupil measurements by 82.5% when the model parameters were pre-set to the physical dimensions of the experimental layout, and by 97.5% when they were optimized to fit the empirical error surface. Thus, our calibration data and correction procedure allow for an unprecedented reduction in PFE without sacrificing data quality or experimental flexibility in the process.

Experiment 1

The *EyeLink 1000 User Manual* (SR Research, 2010, p. 98) states that, “The pupil size data is not calibrated, and the units of pupil measurement will vary with subject setup. Pupil size is an integer number, in arbitrary units. . . . Pupil size measurements are affected by up to 10% by pupil position, due to the optical distortion of the cornea of the eye, and camera-related factors.”

This statement is very vague. It does not even specify the type of scale for these “arbitrary units”. A measurement procedure establishes a correspondence between a set of numbers and a set of objects with respect to some attribute of interest—pupil diameter in our case. It matters greatly what types of relationships are preserved by this mapping (Torgerson, 1958). An *ordinal scale* only preserves comparative relations ($<$, $=$, $>$). An *interval scale* preserves distances but not ratios. Thus, it is meaningful to calculate means, differences, and standard deviations among interval-scale measurements, but they do not warrant inferences of the form, “the pupil diameter in condition A is 10% greater than that in condition B.” The latter kind of statement asserts a multiplicative relationship ($A = 1.1 \times B$) that is warranted only for a *ratio scale*—that is, for an interval scale with a *true origin*. Ratio scales map the number zero to the (possibly hypothetical) object that altogether lacks the attribute in question.

Ideally we want to measure pupil diameter on a ratio

scale with standard units such as millimeters. The next best method is to measure it on a ratio scale whose units are proportional to millimeters, although the coefficient of proportionality is not specified. This is one possible interpretation of the phrase “arbitrary units.” The coefficient of proportionality may vary with experimental setup, as the *Manual* warns, but as long all measurements within a given setup form a ratio scale, they still support multiplicative operations. This is important because, as we show below, the geometric foreshortening law is multiplicative: The foreshortened diameter equals the true diameter *times* the cosine of a certain angle.

The *Manual* provides very little information about how the pupillometric data are acquired and processed by the proprietary software on the host computer. This raises the disconcerting possibility that the “arbitrary units” may only form an interval scale. This could occur, for instance, if the software added an unspecified constant before writing the numbers to the data file. This would render invalid all multiplicative operations with these measurements. In particular, it would undermine the common practice in the literature to report the pupillometric data in terms of percent change from baseline.

Finally, the “arbitrary units” may form merely an ordinal scale. This could happen, for instance, due to “the optical distortion of the cornea . . . and camera-related factors.” In this case, the information content of the pupillometric data would be very low and they should be analyzed using ordinal statistical methods (Agresti, 1984).

The purpose of this preliminary experiment is to collect calibration data to characterize the type of scale used for EyeLink 1000 pupillometry. The results indicate that the “arbitrary units” form a ratio scale with a layout-dependent coefficient of proportionality to millimeters.

Method

We used an EyeLink 1000 desktop eye tracker (SR Research, 2010) at a sampling rate of 1000 Hz. The tracking mode was set to “pupil only,” the pupil threshold parameter to 110, and the eye-to-track parameter to “Left.” Pupil area was measured using centroid mode throughout the study. Centroid mode computes pupil area using a center-of-mass algorithm that identifies the number of black pixels in the thresholded pupil and its center on the camera image. The *User Manual* recommends the center-of-mass algorithm over the ellipse fitting algorithm because “it has very low noise” (SR Research, 2010, p. 71).

A series of 11 black discs with diameters ranging from 2 mm to 7 mm in 0.5 mm increments were printed on white paper using a laser printer. The accuracy of the printed diameter was verified with Neiko digital calipers. An apparatus consisting of a flat piece of wood (157 mm x 88 mm x 36 mm) and an opaque white piece of plastic (50 mm x 95 mm) with a window (13 mm x 13 mm) served to present the printed pupils at a constant position (see Figure 2a). The 7 pupils were printed on a single tape that could be placed underneath the plastic pouch and translated horizontally.

We explored three separate geometric layouts—Near,

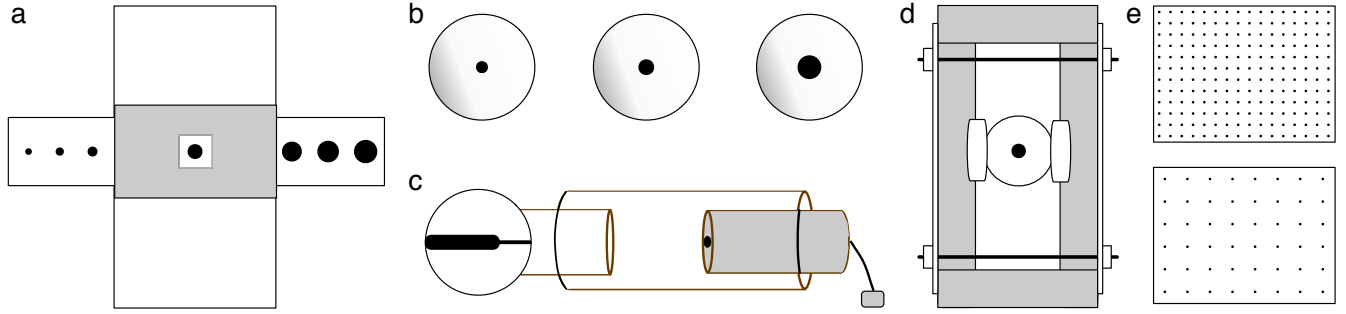


Figure 2. Schematic illustrations of the calibration apparatus, artificial eye sizes, laser orientation apparatus, ocular socket, and target displays. The windowed pupil calibration apparatus (a) allowed steady presentation of each of 7 black dots of known diameter printed on a paper slip. (b) Three spherical artificial eyes were constructed with pupil diameters 3.17 mm, 4.76 mm, and 7.14 mm, respectively. The gaze direction of the artificial eyes were precisely guided using an attached laser pointer (c) whose collimated beam illuminated a bright dot on the target display screen (e). A simple socket (d) allowed the artificial eye to rotate to “fixate” a target while maintaining constant physical position.

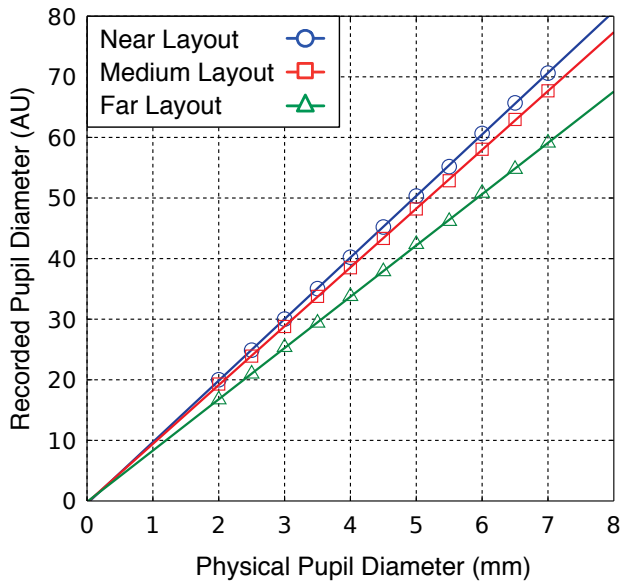


Figure 4. Ratio scale between arbitrary units and physical units. The pupil-diameter data recorded by EyeLink 1000 in arbitrary units are proportional to the true diameter of the physical pupil, but the coefficient of proportionality depends on the distance between the camera and the artificial eye.

Medium, and Far—that varied the relative positions of the camera, artificial eye, and monitor as specified in Figure 3, panels a, b, and c, respectively. At each layout we made pupillometric recordings with each printed pupil for 5 seconds at 1000 Hz, thereby obtaining 5000 pupil-area samples. The area data were converted to linear (diameter) data by taking square roots, and then averaged across the 5000 samples. This procedure produced a calibration data set of 33 pupil diameter measurements (3 layouts \times 11 pupils).

Results and Discussion

Figure 4 plots the calibration data in EyeLink’s “arbitrary units” against the true diameter of the physical pupil in mm. The results clearly indicate that the arbitrary units form a ratio scale within each layout, but the coefficient of proportionality to the true pupil diameter varies across layouts. Specifically, this coefficient was $k_n = 10.07$ for Near, $k_m = 9.65$ for Medium, and $k_f = 8.43$ for the Far layout in Figure 3. Each of these linear regressions accounted for essentially all the variance ($R^2 > .999$) of the 11 measurements within a given layout. Importantly, adding a free intercept parameter to the regression equation did not significantly improve the fits, and the intercept estimates were negligible (< 0.5 AU). This indicates that 0 arbitrary units always correspond to 0 mm regardless of layout, thereby satisfying the true-origin requirement for ratio scales.

Furthermore, the slope coefficients k were themselves approximately inversely proportional to the eye-to-camera distances, which were $L_n = 584$, $L_m = 610$, and $L_f = 698$ mm for the three layouts.¹ To verify this inverse proportionality, we calculated products of the form $p_i = \alpha k_i L_i$, where $i \in \{n, m, f\}$ and $\alpha \approx 1.70 \times 10^{-4}$ is a re-scaling parameter determined in the next paragraphs. If each slope coefficient is inversely proportional to the eye-to-camera distance regardless of layout, then p_i should equal unity for any i . In our data, we obtain $p_n = 0.9999$, $p_m = 1.0002$, and $p_f = 0.9998$.

This suggests that the “arbitrary units” are not really units of length but of visual angle. Apparently, the EyeLink 1000 in centroid mode reports the *angular* area subtended by the pupil. A straightforward algorithm for estimating this angular area—and probably the algorithm implemented by the EyeLink software—is to count the “pupil” pixels in the image of the eye. This interpretation is consistent with the statement in the *User Manual* that the pupil area is recorded in

¹ In Figure 3, these are the hypotenuses of the triangle with vertical side $C_y = 310$ mm giving the height of the eye relative to the camera and with horizontal side $C_z = 495$, 525, and 625 mm, respectively.

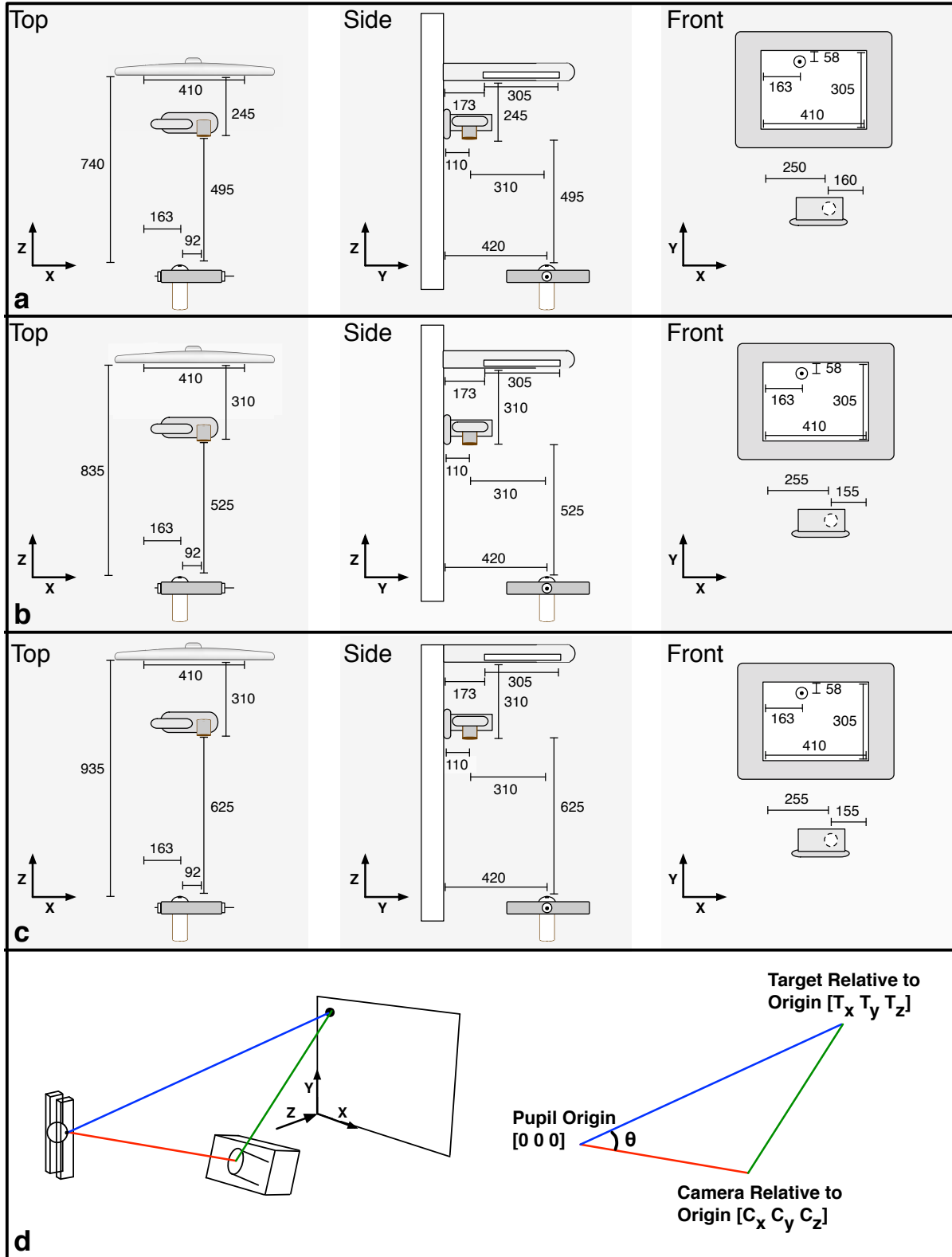


Figure 3. Experimental Layouts and Geometric Model. Three separate geometric layouts (panels a–c) varied the relative positions of the camera, artificial eye, and monitor. Each layout is diagrammed from three vantage points (top, side, and front). All distances are in millimeters. Pupil calibration (Experiment 1) and pupil foreshortening (Experiment 2) were measured for each layout. The geometric model (d) estimates the foreshortening of the pupil area as a function of the cosine of the angle θ between the eye-to-camera axis OC and the eye-to-target axis OT. The origin O of the coordinate system is at the center of the artificial pupil. The X axis is horizontal and parallel to the bottom edge of the screen, growing rightward. The Y axis is vertical, growing upward. The Z axis is perpendicular to the screen, growing inward.

“scaled image pixels” (p. 17). Taking the square root transforms the angular area into a linear visual angle ϕ , which is approximately² proportional to the ratio of the true pupil diameter d and the viewing distance: $\phi \propto d/L$.

On the basis of this invariant relationship, Equation 1 provides a straightforward formula for converting “arbitrary units” to mm at any viewing distance:

$$d \approx \alpha L \phi \quad (1)$$

In this equation, the angle ϕ subtended by the pupil is reported by the EyeLink 1000 in “arbitrary units,” L denotes the eye-to-camera distance in mm, and the re-scaling parameter α is in radians per arbitrary unit. We estimate α from the relationship $d_j/(\phi_{ij}L_i) = \text{const}$ across the 11 pupils j and 3 layouts i . This ratio has a mean of $\alpha = 1.70 \times 10^{-4}$ rad/AU ≈ 35.1 arcsec/AU in our data set (SD = 7.20×10^{-7} rad/AU ≈ 0.149 arcsec/AU). With just one free parameter, Equation 1 accounts for 99.99% of the variance of the 33 calibration measurements.

In conclusion, Experiment 1 established that EyeLink’s “arbitrary units” form a ratio scale with coefficient αL depending on the eye-to-camera distance L for each fixed layout. Also, the pupil-size data seem to track the visual angle subtended by the pupil. Consequently, the units can be converted across layouts according to Equation 1. These results were obtained with very simple artificial pupils printed on a flat surface. The next experiment investigates the pupil foreshortening error caused by the rotation of spherical artificial eyes relative to the camera.

Experiment 2

When an eye is photographed from an oblique angle, the image of the pupil becomes elliptical and the apparent pupil area decreases (e.g., Jay, 1962; Spring & Stiles, 1948). The resulting pupil foreshortening error (PFE) depends on the viewing angle. In a typical eye tracking setup, the camera is stationary relative to the screen but the eye rotates in its socket, thereby varying the angle between the optical axis of the camera and the plane of the pupil. We constructed spherical artificial eyes that could rotate in artificial sockets and could be pointed to “fixate” arbitrary points on the screen. The purpose of Experiment 2 was to map the relationship between EyeLink’s pupillometric data and the screen coordinates of a grid of fixation points. The results indicated that the measurements were contaminated by substantial PFE, but the magnitude of the error depended systematically on the cosine of the angle between the eye-to-camera axis and the eye-to-stimulus axis. Because the PFE was systematic, it could be corrected very well by a simple geometric model.

Method

Apparatus

All EyeLink 1000 settings were the same as those in Experiment 1, with the exception of the pupil threshold parameter. The latter was lowered from 110 to 60 as the spherical

artificial eyes with pupillary wells required a lower threshold than the flat pupils printed on paper. These settings provided a reliable thresholded pupil signal across all measured conditions.

Artificial Eye Model

Three artificial eyes were manufactured from three solid wooden spheres, each 31 mm in diameter. The production of each eye began by drilling a 0.79 mm pilot hole all the way through the center of the wooden ball using a Skil 3320 10” benchtop drill press and DeWalt titanium pilot drill bits. Then a well was drilled 20 mm deep into each eye to form the artificial pupil (see Figure 2b, c). The three eyes had well diameters of 3.17 mm, 4.76 mm, and 7.14 mm, respectively. The inside of each well was painted black and the rest of the eye was painted white using tempera washable paint. Finally, a pressure-switch laser sight (<5 mW) was collimated within a 100-mm section of PVC pipe (25 mm diameter) that was attached to a second 30-mm section of PVC pipe (15 mm diameter) that was firmly cemented to each artificial eye using epoxy cement (Figure 2c). The laser beam originated approximately 75 mm behind the eyeball and traveled to the screen through the .79-mm hole. The pressure switch allowed us to (de)activate the laser without disturbing the eye orientation in the socket.

The artificial eye was kept at a constant position within the headrest using a simple ocular socket mechanism (see Figure 1). This mechanism was composed of 4 pieces of wood (sides 131 x 13 x 28 mm; top and bottom 13 x 63 x 38 mm), 2 steel mending plates (140 x 0.9 x 13 mm), 2 threaded steel rods (152 mm long), 4 wing nuts, and 2 rubber grommets. Two holes were drilled through the wooden sides for the threaded rods—one 14 mm from the top and one 26 mm from the bottom. Two circular recessed holes (25 mm diameter) were drilled 4 mm deep into the inside of each wooden side piece (centered horizontally and vertically) to hold the rubber grommets firmly in place. The ocular socket was then assembled as shown in Figure 2d. The threaded rods were placed through each side piece with a mending plate on each outside edge held in place by the four wing nuts. The wing nuts could then be tightened down to firmly hold the current artificial eye in place and loosened to remove and replace the eye when needed. The elasticity of the rubber grommets allowed us to manipulate the elevation and azimuth of the artificial eye within the socket, while at the same time holding firmly when the desired orientation was reached.

The standard EyeLink 1000 nine-point calibration and validation procedures were used to quantify the accuracy of the artificial eye and socket apparatus. The nine-point validation data indicated high gaze-position accuracy across the display (average mean error $M=0.18$, $SD=0.04$; maximum error $M=0.26$, $SD=0.04$ degrees of visual angle). The validation data verified that the laser pointers were collimated

² We use the paraxial approximation $\tan \phi \approx \sin \phi \approx \phi$ for small angles measured in radians (Atchison & Smith, 2000).

accurately and the ocular socket kept the artificial eyes in a stable orientation throughout the procedure.

Gaze Position Map and Procedure

Pupil size measurements were collected with each artificial eye on a rectangular grid of fixation targets for each experimental layout. For the Near and Far layouts, we used $16 \times 12 = 192$ targets spaced 64 pixels apart (see Figure 2e top); for the Medium layout, we used $8 \times 6 = 48$ targets spaced 128 pixels apart (at screen resolution 1024×768).

On each trial, we rotated the eye until the pointer beam illuminated the appropriate target on the screen. A box shielded the eye tracking camera as a precaution against accidental exposure to the beam. As in Experiment 1, pupil-area measurements were recorded for 5 s at 1000 Hz, converted to linear units by taking square roots, and averaged across the 5000 samples. This produced one pupil-diameter datum per trial.

Geometric Model

The results from Experiment 1 suggested that the EyeLink 1000 (operating in centroid mode) reports the *angular* area subtended by the pupil. Based on simple geometric principles, Equation 1 accounted extremely well for the pupillometric data when the distance between the camera and the eye was varied. Our working hypothesis is that the same principles can account for the pupillometric data when the *angle* between the camera axis and the eye axis is varied. In this section, we develop a geometric model that formalizes this hypothesis and provides explicit formulas for correcting the pupil foreshortening error.

The key term in the model is the *oblique angle* θ between the eye-to-camera axis and the eye-to-stimulus axis (Figure 3d). To simplify the trigonometric calculations, the model assumes the camera is pointed directly at the eye and thus the eye-to-camera axis coincides with the optical axis of the camera. Let us consider first the *baseline configuration* in which the eye too is pointed directly at the camera. Then the eye axis coincides with the camera axis and $\theta = 0$. Let A_0 denote the angular area subtended by the pupil as viewed from the camera in the baseline configuration.

Our goal is to quantify the foreshortening effect, whereby the apparent angular area $A(x, y)$ diminishes when the eye rotates away from the camera to fixate a target with coordinates x and y on the screen. The derivation proceeds in two steps: First, we argue that the foreshortening is multiplicative and depends on the oblique angle θ according to Equation 2. Second, we express $\cos \theta$ as a function of x and y using simple vector calculus.

$$A(x, y) = A(\theta(x, y)) = A_0 \cos \theta(x, y) \quad (2)$$

We consider a circular pupil for concreteness, although the mathematical derivation generalizes to arbitrary planar shapes. When a circular disk is viewed from an oblique angle, it is projected into an ellipse and its apparent area decreases (Gagl et al., 2011; Jay, 1962; Mathur, Gehrmann, &

Atchison, 2013; Spring & Stiles, 1948). Importantly, the projected pupil becomes narrower in the direction of view (the *tangential section*) but remains unchanged in the perpendicular direction (the *sagittal section*, Equation 3.2 in Atchison & Smith, 2000). This is why the multiplier in Equation 2 above is $\cos \theta$ rather than $\cos^2 \theta$. To simplify the analysis, we have assumed that the eyeball diameter is negligible relative to the distance to the camera and, consequently, the rotational center of the eyeball lies approximately in the pupil plane. This seems a reasonable assumption for desktop trackers such as the one used in this article, but probably becomes very crude for head-mounted trackers. Taking the eyeball geometry into account does not change any of the principles here, but complicates the trigonometry. Another assumption that is implicit in Equation 2 is that the camera has negligible optical aberrations near the center of its field of view.

The second step of the derivation is to express $\cos \theta$ as a function of the target coordinates x and y . It is convenient to work in a coordinate system with an origin O at the pupil center and axes as described in the caption of Figure 3d. The camera lens is at point C with coordinates C_x , C_y , and C_z , which are parameters of the physical layout (Figure 3). The other parameters of the layout are the coordinates S_x , S_y , and S_z of the upper left-hand corner of the screen. Then a fixation target T with screen coordinates x and y mm has coordinates $T_x = x - S_x$, $T_y = y - S_y$, and $T_z = S_z$ in the eye-centered system. In this notation, the oblique angle θ is the angle COT in Figure 3d and its cosine can be calculated via the dot product of the vectors OC and OT :

$$\cos \theta(x, y) = \frac{C_x T_x + C_y T_y + C_z T_z}{\sqrt{C_x^2 + C_y^2 + C_z^2} \sqrt{T_x^2 + T_y^2 + T_z^2}} \quad (3)$$

For ease of comparison to the physical pupil diameter of the artificial eye models, the angular area A is converted to arc length ϕ using the relationship $\phi = 2 \sqrt{A/\pi}$. Taking square roots of both sides in Equation 2 and rearranging leads to Equation 4, in which ϕ_0 denotes the angle subtended by the pupil diameter in the baseline configuration and $\phi(x, y)$ denotes the apparent angle when the artificial eye points at a target with screen coordinates x and y .

$$\phi_0 = \frac{\phi(x, y)}{\sqrt{\cos \theta(x, y)}} = \text{const for all } x, y. \quad (4)$$

The true pupil diameter d can be determined by pointing the artificial eye directly at the camera (or asking the human participant to look at the camera), obtaining EyeLink measurements of the subtended angle ϕ_0 in “arbitrary units,” and converting them to millimeters according to Equation 1. The pupil foreshortening error $PFE(x, y) = \phi(x, y)/\phi_0$ at fixation point (x, y) is $\sqrt{\cos \theta(x, y)}$. For an artificial eye with a constant pupil, the geometric model predicts that the ratios in Equation 4 will be invariant for all points (x, y) . This prediction is tested below.

Two versions of the geometric model were compared: a parameter-free model and an optimized model. The

parameter-free model simply calculated $\cos \theta$ according to Equation 3 using the physical measurements taken directly from each experimental layout (Figure 3). The relevant measurements are the camera lens coordinates (C_x , C_y , C_z) and the coordinates (S_x , S_y , S_z) of the upper left-hand corner of the screen for the respective layout. The parameter-free model is an idealization that does not take into account the error of the layout measurements or the optics of the camera lens itself, which can alter the effective geometric layout. In order to account for these additional sources of error, an optimized model was also considered. The optimized model fit 5 parameters C_x , C_y , S_x , S_y , and S_z ; the C_z parameter was fixed to the respective physical layout measurement. The optimized model was fit to the pupillometric data separately for each layout using an unconstrained Nelder-Mead optimization routine that minimized the root mean squared error. The initial values for the iterative parameter search were set to the physical layout measurements.

Results and Discussion

Pupil measurements for the 3 spherical artificial pupil diameters (3.17 mm, 4.76 mm, 7.14 mm) and 3 experimental layouts (Near, Medium, Far) resulted in a total of 9 empirical maps. Figure 5 shows the measured pupil diameter as a function of visual angle for each unique map. Pupil diameter is expressed on a common scale across all 9 maps as the deviation from the map-specific geometric mean. Despite the fixed pupil diameter of the artificial eye models, the average range in deviation from the geometric mean for the Near, Medium, and Far Layouts were 14.4%, 10.1%, and 8.4% respectively (see Table 1). Pupil diameter systematically decreased as the eye rotated further away from the camera lens—producing larger variability in pupil diameter as viewing distance decreased and visual angles to the target grid points on screen increased. These results establish the existence of large PFE as a result of changes in gaze position for each measured combination of pupil diameter and experimental layout.

PFE was invariant across the 3 different artificial eye diameters within a given experimental layout. Principal-component analysis (PCA, Everitt & Dunn, 2001) was performed to estimate the intra-map similarity across the 3 pupil diameter maps within the Near, Medium, and Far Layouts. The first principal component accounted for nearly all the variance across the three pupil diameter maps (Near 99.8%; Medium 98.9%; Far 98.9%). The invariance across changes in pupil diameter is a critical finding as it establishes that the PFE surface is not affected by the changes in pupil diameter that occur when measuring a dynamic, biological eye. Due to the invariance across pupil diameter the maps were collapsed by taking the geometric mean across the 3 different pupil diameter maps and normalizing relative to the geometric mean within a given experimental layout, resulting in one aggregate map for each experimental layout. The remaining results focus on modeling and correcting PFE in the aggregate layout maps.

We evaluated the validity of the geometric model in Equations 3 and 4, as well as its ability to correct the PFE and

Table 1

Descriptive statistics of the deviation in pupil diameter from the geometric mean for the 9 empirical maps (3 experimental layouts \times 3 pupil sizes). RMSE = root mean squared error.

	RMSE	Min	Max
<i>Near Layout</i>			
Pupil Diameter 3.17	0.034	0.916	1.058
Pupil Diameter 4.76	0.032	0.918	1.053
Pupil Diameter 7.14	0.038	0.905	1.060
<i>Medium Layout</i>			
Pupil Diameter 3.17	0.025	0.940	1.044
Pupil Diameter 4.76	0.026	0.938	1.043
Pupil Diameter 7.14	0.024	0.944	1.037
<i>Far Layout</i>			
Pupil Diameter 3.17	0.022	0.945	1.039
Pupil Diameter 4.76	0.019	0.951	1.030
Pupil Diameter 7.14	0.019	0.951	1.030

recover the true pupil diameter ϕ_0 . As a result of the normalization and aggregation described above, $\phi_0 = 1 = \text{const}$ for each map and all residual variation is due to foreshortening error. The parameter-free model reduced the root mean squared error of the aggregated measurements by an average of 82.5%. The correction is applied according to Equation 4 by dividing the measured pupil diameter at each grid target location by the square root of the cosine of its corresponding angle. Figure 6 shows the empirical aggregate layout maps before correction, the geometric model corrective multipliers, and the corrected maps. There was a strong correspondence between the 3 aggregate maps and their respective multipliers (Near $R^2=.99$, Medium $R^2=.98$, Far $R^2=.98$). The parameter-free geometric correction greatly reduced the RMSE to 18.7%, 15.8%, and 18.2% of its original value for the Near, Medium, and Far layout. Table 2 shows the descriptive statistics before and after applying the parameter-free correction.

While the parameter-free geometric model reduced the RMSE in pupil diameter substantially, it left a small amount of residual error that was systematic and thus could be corrected further. Specifically, close inspection of the third column in Figure 6 suggests that the correction generated by the parameter-free model was not strong enough to fully correct the “tilt” of the uncorrected maps in the first column. Thus, even after the parameter-free correction, the pupil diameter remained slightly underestimated for fixation targets near the left edge of the screen and slightly overestimated for targets near the right edge.

To evaluate the extent to which this residual error can be corrected still further, the parameters of the geometric model were optimized to fit each aggregate map. Figure 7 shows the three empirical maps, the optimized corrective multipliers, and the optimized correction maps. The RMSE was reduced to 2.0%, 2.5% and 3.2% of its original value for the Near, Medium, and Far layout, respectively, using the optimized parameters listed in Table 3. As can be seen in Table 2, the

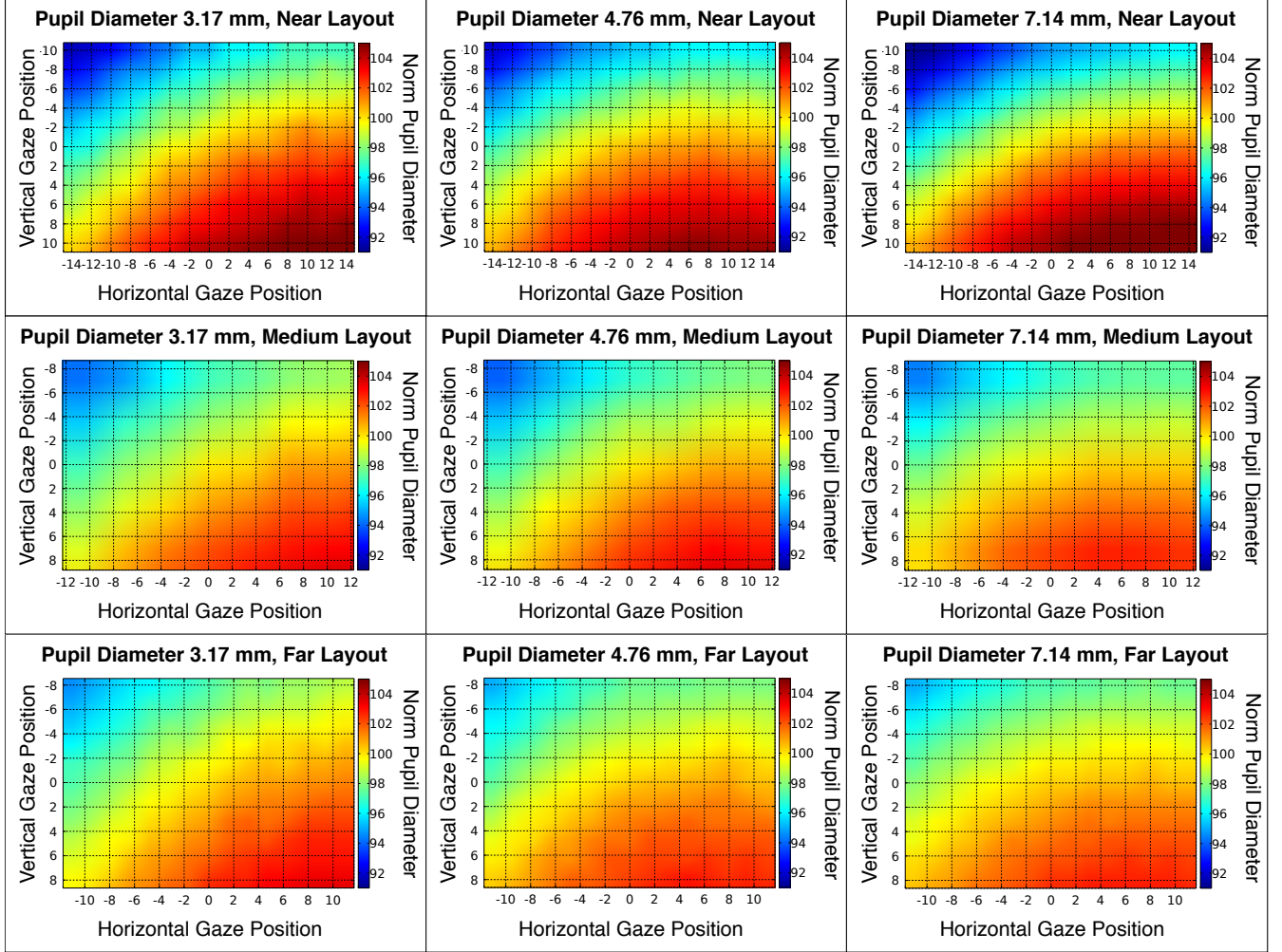


Figure 5. Deviation in pupil diameter by pupil size and layout. Each panel shows the deviation in measured pupil diameter from the geometric mean for a unique combination of pupil size (3.17 mm, 4.76 mm, 7.14 mm) and experimental layout (Near, Medium, Far). Large deviations in measured pupil diameter were observed across each map despite the fixed pupil of each artificial eye. Deviations in measured pupil size grew as layout distance became shorter and visual angles increased. Deviations were relatively invariant across pupil diameter within a given experimental layout. Gaze positions are in degrees of visual angle.

deviation from the geometric mean was less than 0.1% across the three optimized correction maps, which is well within the uncertainty of measurement of a *stationary* pupil across trials. This indicates that the PFE can be eliminated entirely for practical purposes, at least for the artificial eyes considered here, provided enough calibration measurements are available to constrain the parameters of the geometric model.

It is instructive to compare the optimized parameter values in Table 3 to their physical counterparts. The distance S_z from the eye to the screen was constrained well by the data but there were surprisingly large deviations in the other four parameters. The optimized values of C_x and S_x would be veridical if the artificial eye were located ≈ 60 mm closer to the left edge of the screen (on average across the three layouts). The optimized values of C_y and S_y would be veridical if the eye were located ≈ 90 mm lower. The overall effect

was to produce greater variation in the oblique angle θ in Equation 3 across each map, which in turn produced stronger correction multipliers. The within-map RMSE of the optimized model multipliers were larger than their parameter-free counterparts and better matched the RMSE of the calibration data (Table 2). We speculate that the optimized parameters captured the magnification effect of the optics inside the eye-tracking camera. Equations 3 and 4 essentially model a pinhole camera, whereas the real device includes lenses on the optical path. This is a topic for further research. The theoretical significance of the optimized model fit is that the PFE surfaces were smooth and could be parameterized with a few numbers that could be estimated from calibration data.

Finally, it is useful to evaluate whether binocular pupil recording can alter these conclusions. In the absence of binocular calibration data, we rely on predictions from the

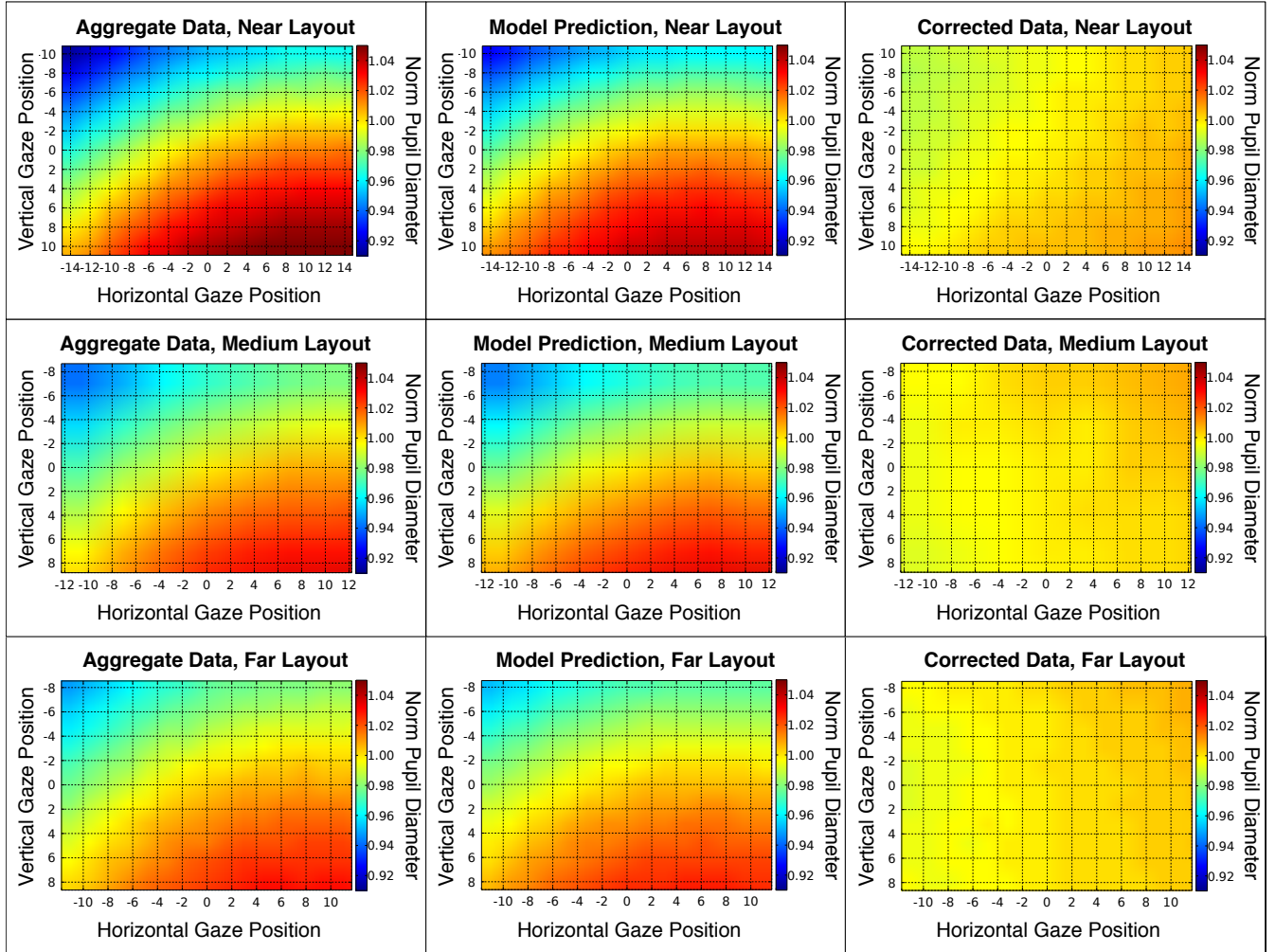


Figure 6. Aggregate empirical maps, parameter-free geometric model multipliers, and parameter-free correction. Each panel shows the deviation in pupil diameter from the geometric mean. The first column of panels shows the aggregate empirical maps collapsed across pupil size for each experimental layout. The second column shows the corrective multipliers produced by the parameter-free geometric model. The third column shows the corrected data produced by dividing the aggregate empirical data by its respective corrective multiplier. The parameter-free geometric correction reduced the within-map RMSE to 18.7%, 15.8%, and 18.2% of its original value for the Near, Medium, and Far layout. Gaze positions are in degrees of visual angle.

parameter-free geometric model to estimate the extent to which the PFE for the two eyes would cancel out in a binocular average. All calibration measurements and model calculations thus far used monocular recording from the left-eye position. To simulate the corresponding right-eye position, it is sufficient to subtract the interpupillary distance from the C_x and S_x model parameters (Figure 3). A recent survey of 3976 adults (Dodgson, 2004) reports a median interpupillary distance of 63 mm. We thus assume $C_{x,right} = C_{x,left} - 63$ and $S_{x,right} = S_{x,left} - 63$. Table 4 summarizes the predicted PFE for the left and right monocular cases and their binocular average. Binocular pupil recording would have decreased the pupil foreshortening RMSE only by about 1% relative to the monocular recording from the left eye. It should be noted that if we had used the right eye position, binocular aver-

aging would have *increased* the RMSE of the PFE by 1%. This is because there is less PFE for whichever eye is closer to the camera, all else being equal. In sum, the geometric model predicts that binocular data acquisition would have a negligible impact on the PFE.

General Discussion

This article reports two experiments that investigate and calibrate the pupillometric measurements of a tabletop EyeLink 1000 eye tracker (SR Research, 2010) using artificial eyes with known pupil diameter. Experiment 1 established that EyeLink’s “arbitrary units” form a ratio scale with a true zero. This result justifies the common practice in the field to express *relative* pupil diameters as dimensionless *ratios*—e.g., percent change from baseline. When absolute measure-

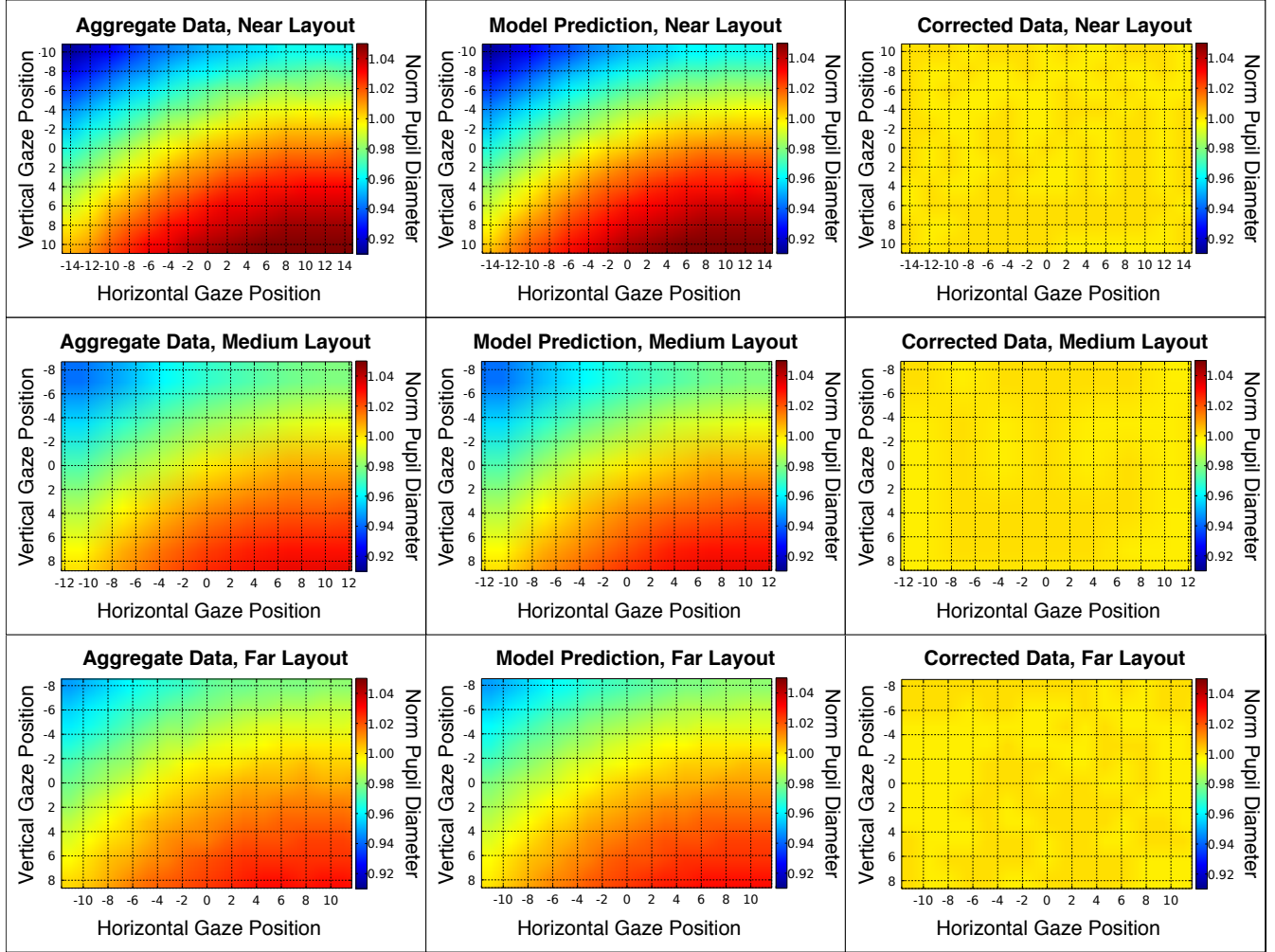


Figure 7. Aggregate empirical maps, optimized geometric model multipliers, and optimized correction. Each panel shows the deviation in pupil diameter from the geometric mean. The first column of panels shows the aggregate empirical maps collapsed across pupil size for each experimental layout. The second column shows the corrective multipliers produced by the geometric model with optimized parameters. The third column shows the corrected data produced by dividing the aggregate empirical data by its respective corrective multiplier. The optimized geometric correction virtually eliminated pupil foreshortening error, reducing the within-map RMSE to 2.0%, 2.5%, and 3.2% of its original value for the Near, Medium, and Far layout. Gaze positions are in degrees of visual angle.

ments are needed, they can be calculated for a given arrangement of the eye tracking camera, headrest, and monitor: The “arbitrary units” are proportional to millimeters but the coefficient of proportionality depends on the eye-to-camera distance (Equation 1). Furthermore, the results of Experiment 1 strongly suggest that EyeLink’s pupil-size data are expressed not in units of linear length but of the angle subtended by the pupil as viewed from the camera. These units can be converted across layouts according to Equation 1.

Experiment 2 mapped the pupil foreshortening error (PFE) when the artificial eye rotated in its socket to fixate a grid of target points on the screen. Data were collected across 3 experimental layouts with spherical artificial eyes with 3 pupil diameters. The 9 resulting maps showed large PFE that increased as a monotonic function of the *oblique angle*

between the eye-to-camera axis and the eye-to-target axis. The results supported three major conclusions as follows: First, the *relative* PFE maps were highly replicable across different pupil sizes, as evident in the high ($R^2 > .98$) correlation among maps collected with different artificial eyes in a given experimental layout. This suggests that the relative PFE varies only as a function of the orientation of the eye with respect to the camera but does not depend on the pupil diameter. In other words, the eye-tracking hardware and software operate linearly with respect to pupil diameter. Second, a parameter-free geometric model reduced the error to <18% of its uncorrected value. This suggests that the PFE is largely due to geometric foreshortening of the type captured by Equations 3 and 4. Note that the foreshortening follows a multiplicative law (Equation 4). The third major

Table 2

Descriptive statistics of the deviation in pupil diameter from the geometric mean for the aggregate layout maps, the corrective multipliers generated by the parameter-free (PF) model and the model with optimized (opt.) parameters, and the corrected data (parameter-free and optimized). RMSE = root mean squared error.

	RMSE	Min	Max
<i>Near Layout</i>			
Aggregate data	0.034	0.913	1.056
Model multipliers (PF)	0.029	0.925	1.046
Corrected data (PF)	0.006	0.987	1.014
Model multipliers (opt.)	0.034	0.912	1.056
Corrected data (opt.)	<0.001	0.998	1.002
<i>Medium Layout</i>			
Aggregate data	0.025	0.941	1.040
Model multipliers (PF)	0.025	0.944	1.038
Corrected data (PF)	0.004	0.993	1.009
Model multipliers (opt.)	0.025	0.941	1.039
Corrected data (opt.)	<0.001	0.998	1.002
<i>Far Layout</i>			
Aggregate data	0.020	0.949	1.033
Model multipliers (PF)	0.019	0.952	1.030
Corrected data (PF)	0.004	0.992	1.008
Model multipliers (opt.)	0.020	0.947	1.033
Corrected data (opt.)	<0.001	0.998	1.002

conclusion of Experiment 2 was that model-based optimization could reduce the error to levels comparable to the measurement precision for a stationary pupil. In other words, the foreshortening error induced by the rotation of the eye could be eliminated almost entirely. Such accurate correction is possible because the PFE surface is smooth and can be described with five parameters that can be estimated from calibration data, at least for artificial eyes.

On the basis of these empirical results we formulate three theoretical principles. They seem obvious in hindsight but were neglected in earlier PFE research (e.g., Brisson et al., 2013; Gagl et al., 2011). First, PFE analysis must be based on the three-way relationship between the eye, the fixation point on the screen, and the eye-tracking camera. Leaving the camera out of the equation leads to an ill-posed problem that does not permit a general solution. This is revealed when one compares data from different experimental layouts as was done here. Second, the foreshortening law is multiplicative in nature (Equation 2). Therefore, division rather than subtraction is the appropriate corrective operation (Equation 4), and the geometric rather than arithmetic mean is the appropriate measure of central tendency. Additive operations are justified only approximately for normalized data on the basis of the approximation $\log(1+x) \approx x$ for small x . The third principle is the cosine relationship embedded in the geometric model (Equations 3 and 4) that grounds the PFE in the physics of the data acquisition process. With parameters that could have been fixed before any calibration data were col-

Table 3

Comparison of geometric models (parameter-free physical measurements vs. optimized parameters) for Near, Medium, and Far Layouts. All values are distances in mm. C_x , C_y , C_z are the coordinates of the camera; S_x , S_y , S_z are the coordinates of the upper left-hand corner of the screen (cf. Figure 3). An asterisk indicates the parameter was fixed to the physical layout measurement.

Model Parameters	Near	Medium	Far
<i>Camera-to-pupil, C_x</i>			
Physical	92	92	92
Optimized	130	165	183
<i>Camera-to-pupil, C_y</i>			
Physical	-310	-310	-310
Optimized	-215	-239	-230
<i>Camera-to-pupil, C_z</i>			
Physical	495	525	625
Optimized	495*	525*	625*
<i>Screen-NW-corner, S_x</i>			
Physical	-163	-163	-163
Optimized	-142	-87	-76
<i>Screen-NW-corner, S_y</i>			
Physical	58	58	58
Optimized	206	140	156
<i>Screen-NW-corner, S_z</i>			
Physical	740	835	935
Optimized	736	851	937

lected, the geometric model accounted for over 82% of the variance in these data. In agreement with the first two principles, the model specifies a three-way multiplicative relationship between the coordinates of the eye, fixation target, and camera. The same principles are widely used in computer vision (Forsyth & Ponce, 2011).

Comparison with Previous Research

Our PFE measurements are broadly consistent with published results that measured a single horizontal scan line using an artificial eye (Gagl et al., 2011), and a circular object pursuit task performed by human participants (Brisson et al., 2013). During the object pursuit task of Brisson et al. (2013), the pupil diameter systematically decreased as the object receded vertically and horizontally from the camera. The geometric model is consistent with this pattern of results. Gagl et al. (2011) reported a piecewise linear and quadratic function in their artificial eye measurements. Whereas our PFE maps exhibited a smooth quadratic trend, Gagl et al. (2011) reported that their artificial eye measurements showed a sharp linear increase near the left edge of the screen that then decreased quadratically as the gaze position moved rightward across the display. The quadratic pattern is consistent with the geometric model and predicts that the camera lens was on the left side of their EyeLink 1000 since the right eye position was measured. We applied our Equations 3 and 4 to

Table 4

Predicted deviation in pupil diameter from the geometric mean for each experimental layout for monocular (either left or right eye only) and binocular recording. The corrective multipliers were generated by the parameter-free (PF) model assuming an interpupillary distance of 63 mm. RMSE = root mean squared error.

	RMSE	Min	Max
<i>Near Layout</i>			
Left eye (PF)	0.0294	0.9247	1.0463
Right eye (PF)	0.0289	0.9305	1.0453
Binocular (PF)	0.0291	0.9276	1.0457
<i>Medium Layout</i>			
Left eye (PF)	0.0251	0.9436	1.0381
Right eye (PF)	0.0246	0.9484	1.0368
Binocular (PF)	0.0248	0.9460	1.0374
<i>Far Layout</i>			
Left eye (PF)	0.0190	0.9519	1.0303
Right eye (PF)	0.0186	0.9554	1.0294
Binocular (PF)	0.0188	0.9536	1.0298

a reconstruction of the layout of Gagl et al. (2011) based on their Method description and the EyeLink specifications (SR Research, 2010, p. 7). The geometric model reproduced the qualitative nonmonotonic shape of their Figure 1, although our reconstruction located the peak close to the center of the display rather than at the left edge. The abrupt drop in recorded pupil size near the left edge is anomalous and does not fit the geometric-model predictions.

One plausible explanation for the different PFE function obtained by Gagl et al. (2011) is that they used a tower mount EyeLink 1000, whereas we used a desktop EyeLink 1000. Unlike the latter, the tower mount system does not record a direct image of the eye. Instead, the camera is mounted above the observer and records a reflected image of the eye from an angled infrared hot mirror (SR Research, 2010, p. 7).³ The infrared mirror reflects infrared light but is transparent to visible light, allowing participants to perform tasks on screen that require them to reach in front of them without occluding the camera's image of the eye. The data of Gagl et al. (2011) suggest that the addition of this angled infrared mirror alters the PFE surface in ways that are not well captured by our simple geometric model. The presence of an adjustable mirror violates the model's assumption that the eye-to-camera axis coincides with the optical axis of the camera. The unknown optical properties of the mirror introduce additional complications. The geometric model is best suited to eye tracking systems in which the camera records a direct image of the eye. Researchers that use eye tracking systems with intermediary mirrors are advised to map the PFE surface of their mirrored system explicitly using an artificial eye. The model probably can be modified to account for the mirror by "unfolding" the optical path, but this should be tested on explicit calibration data.

Recommendations

One key objective of the current article is to provide the research community with a technique to correct for PFE in a manner that does not sacrifice experimental flexibility. The present methodology allows for unprecedentedly accurate PFE correction while preserving the freedom to study tasks such as reading or visual search that involve free viewing of the display. Both the parameter-free and the optimized geometric model reduced PFE enough that cognitive effects >2% should be detectable. One key advantage of the parameter-free geometric correction is that it does not require the researcher to collect any artificial-eye measurements. One weakness is that it does not account for the effects of the camera optics that alter the effective geometric layout. The optimized model virtually eliminates PFE by accounting for all sources of error, but requires calibration data to constrain the model parameters. For researchers that wish to use the optimized geometric model and use the EyeLink 1000, Table 3 lists the best-fitting parameters for the 3 layouts in our study. For researchers that use other eye trackers and/or layouts, the detailed description of our artificial eye and socket apparatus can help them calibrate their respective systems. Desktop or monitor mounted trackers (e.g., Tobii T60 and T120, Tobii Technology, 2011; Tobii X60 and X120, Tobii Technology, 2008; SMI RED series, SensoMotoric Instruments, 2009) that record direct images of the eye are more likely to correspond well to the geometric model. Eye tracking systems that do not directly image the eye but rely on reflected infrared images of the eye (e.g., tower mount EyeLink 1000 and SMI iView X Hi-Speed, SensoMotoric Instruments, 2009) are more likely to deviate from the geometric model due to the additional optics of the intermediary infrared mirror (e.g., Gagl et al. (2011)). The geometric relationships formalized in Equations 2–4 apply to all camera-based trackers, but the proprietary post-processing software of specific manufacturers may render these equations unsuitable for PFE correction purposes.⁴ Finally, some of the simplifying assumptions used in the derivation of these equations will have to be re-examined for head-mounted trackers that put the camera much closer to the eye and thus the diameter of the eyeball is no longer negligible. These assumptions were made explicit in the Model section above.

The magnitude of the pupil foreshortening error depends greatly on the geometric configuration of the screen and camera relative to the eye (Figure 3). The following general recommendations can be used to inform the choice of layout. Putting the display far from the participant tends to reduce the range of oblique angles θ and hence the PFE. On the other

³ The typical camera-to-eye distance in the tower mount system is 380 mm, which is comparable to our Near Layout (425 mm).

⁴ We urge the eye tracking manufacturers to provide clearer descriptions of the broad type of algorithm used to estimate pupil size (e.g., pixel counting vs. ellipse fitting) and the scale type (e.g., interval vs. ratio) of the resulting measurements. We understand the motivation to keep the detailed algorithms proprietary, but the manual should provide enough information to allow the user community to interpret the data.

hand, it increases the error in estimating the fixation location. This creates a trade-off between pupillometric accuracy and eye-tracking accuracy because the latter deteriorates when the screen is far from the participant. An elegant solution is possible when the research question requires that the stimuli appear at multiple locations but does not constrain the spatial relationship among these locations. Our recommendation for such situations is to spread the stimuli along an arc that maintains a constant oblique angle θ with the camera-to-eye axis. The coordinates of these locations can be calculated from Equation 3 under the constraint $\cos \theta(x, y) = k = \text{const}$ for a suitably chosen k . The geometric model predicts identical PFE for all points along an arc of this type. Analogous circular contours of equal PFE are prominent in the calibration data (Figure 5). Another general recommendation is to place the camera as close to the eye-to-stimulus line of sight as the camera dimensions allow. Such placement makes $\cos \theta$ as close to 1 as possible. Recording monocularly from the eye closest to the camera is predicted to provide modest ($\approx 2\%$) reductions in PFE relative to the more distant eye.

To facilitate effective PFE correction, the following best practices are recommended for the data collection phase: First, we strongly discourage the participation of observers whose eyelids partially occlude their pupils as well as participants that wear glasses or contact lenses as these can each affect the PFE in unpredictable ways. For desktop or screen mounted eye tracking systems, a chin-and-forehead rest should be used to keep the eye in a constant position. When possible, the rest height should be kept constant across participants. In our lab we adjust the height of the chair to accommodate participants of different stature, leaving the chin-and-forehead rest fixed at all times. Thus, a common geometric model can be applied to the entire data set during post-processing. If the chin-and-forehead rest height must be adjusted on an individual basis, height measurements should be logged for each participant and an individualized geometric model should be applied. Finally, regardless of whether the proprietary image-based eye tracking software uses pixel counting or ellipse fitting algorithms to estimate pupil size, the pupil threshold parameter should be set at the beginning of the experimental session and not altered during the course of the session. Changing the threshold parameter in mid-session will result in large changes in recorded pupil size.

Last but not least, we offer guidelines for evaluating published pupillometric data for possible PFE artifacts. The main danger is the possibility for *systematic* pupil foreshortening errors. Nonsystematic errors merely increase the measurement noise and cannot alter the substantive conclusions of an experiment, even though they are still undesirable because they can cause type II errors. Systematic PFE, however, is very dangerous because it can cause type I errors and cannot be mitigated through averaging. For example, a systematic PFE confound would arise if all emotional stimuli were presented at one location, whereas all neutral stimuli at another. Such obvious cases of bad experimental design would be screened out during the peer review process, but there are subtler cases that are very easy to miss. Consider the following example based on an experiment from our lab.

We presented all stimuli at a single location and analyzed only trials in which the participants maintained fixation on the stimulus, as instructed. Following the widespread practice in the literature, we believed that we had thereby eliminated any possibility for PFE. We were wrong, due to a subtle interaction between certain features of our experimental procedure and certain regularities in the participants' behavior. It is instructive to consider this case in some detail. Our procedure included a free viewing period from the stimulus offset until the end of the trial and through the inter-trial interval. Such free viewing periods are very common in eye-tracking research because it is very uncomfortable to maintain constant fixation for an extended period of time. Also, mindful of the time lag between the stimulus and the task-evoked pupillary response (Partala & Surakka, 2003; Murphy et al., 2011; Hayes & Petrov, 2015), we analyzed the pupillary data from the entire trial, including the free period. This is also standard practice, but it allowed systematic PFE to creep in. The problem was that the participants' pattern of free viewing depended systematically on the preceding stimulus. After an easy stimulus, their eyes roamed the screen more or less at random. After a difficult stimulus, however, the participants often checked the bonus score that was displayed at the top of the screen at all times. Apparently, they were less confident in their responses on difficult trials and tended to look at the accuracy-contingent score to reassure themselves. This behavior produced a preponderance of fixations near the top of the screen on difficult trials, which in turn introduced a systematic PFE masquerading as a cognitive-load effect in (apparent) pupil diameter a few hundreds of milliseconds after the stimulus offset. This example illustrates that *systematic* PFE confounds can arise through subtle interactions among seemingly unrelated features of the experimental setup. Free viewing periods are particularly vulnerable to PFE artifacts because the participants can introduce unexpected regularities through their patterns of free looking. Therefore, presenting the stimuli at a constant location (or at counterbalanced locations) is guaranteed to protect against PFE only if all subsequent pupillometric analyses are based entirely on data collected while fixation was maintained within the controlled location(s).

Note that we do not claim that all published pupillometric research is infested with PFE artifacts and is thereby useless. The constant-fixation and counterbalancing methods are widely used and do mitigate the PFE, although the example above shows that they are not as foolproof as it is commonly assumed. The stimuli in many experiments are presented near the center of the screen where the PFE is smaller compared to the far periphery. Even imperfect measurements can provide a wealth of evidence that is relevant and replicable. Still, at best the PFE erodes the statistical power and complicates the interpretation, and at worst it can introduce spurious regularities and/or change the magnitude of the real effects. The main purpose of this article is not to throw a dark shadow on the pupillometry literature, but to provide a method for correcting the PFE and encourage other researchers to adopt it in their future work.

Limitations

One potential limitation of our study is that we do not explicitly model the refracting effect of the cornea—the transparent curved layers in front of the iris (Atchison & Smith, 2000). The image recorded by video-based eye trackers measures the *entrance pupil*—the aperture stop of the iris as viewed through the refractive surfaces of the cornea. The refractive power of the cornea is considerable; it is approximately two times stronger than that of the lens of a relaxed eye (Atchison & Smith, 2000). The absolute magnification provided by the cornea does not affect the relative PFE. Rather, our concern is about corneal *anisotropy*. If different segments of the cornea have different refractive indices, the pupil diameter can appear to change when the eye rotates, turning different segments towards the camera. There are empirical data that bear on this issue (e.g., Spring & Stiles, 1948; Jay, 1962; Jennings & Charman, 1978). These studies photograph human eyes from a range of oblique angles, fit ellipses to the pupil images, and estimate the foreshortening coefficient as a function of the angle. A recent article (Mathur et al., 2013) re-analyzed the data from six published studies plus sophisticated new measurements. The overall conclusion was that, “Off-axis pupil shape is well described by a cosine function that is both decentered by a few degrees [towards the temporal side] and flatter by about 12% than the cosine of the viewing angle.” (Mathur et al., 2013, p. 7).

The empirical foreshortening function of Mathur et al. (2013) can easily be incorporated into our geometric model. To that end, we substitute their Equation 9 for the term $\cos \theta$ in our Equation 4, yielding:

$$\phi_0 = \frac{\phi(x, y)}{\sqrt{0.992 \cos\left(\frac{\theta(x, y) + 5.3}{1.121}\right)}}, \quad (5)$$

where $\theta(x, y)$ is determined from Equation 3, in degrees. Note that the data of Mathur et al. (2013), as well as the earlier data re-analyzed therein (Spring & Stiles, 1948; Jay, 1962; Jennings & Charman, 1978, etc.), were all collected along the horizontal equatorial line. One area for future research would be to systematically map the apparent pupil foreshortening across the entire visual field in participants with dilated eyes, and compare the human eye data to our artificial eye data to quantify the effect of corneal refraction.

Conclusions

Pupil foreshortening error is a potentially large confound that should be taken into account prior to interpreting pupillary data from image-based eye trackers. We introduced an artificial eye and socket model and systematically mapped the PFE across the horizontal and vertical visual field in three separate experimental layouts. In light of previous cognitive pupillometric research, our data indicate that the PFE can be larger than many cognitive pupillometric effects. The PFE is not currently corrected by popular commercial image-based eye tracking systems. We formulated a simple parameter-free geometrical model that reduced the deviation of the PFE

by 82.5% and an optimized model that reduced it by 97.5%. Thus, very accurate PFE correction is possible and the corrected pupillometric data have the precision necessary to measure typical cognitive effects without limiting the types of tasks that can be studied by researchers.

References

- Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: John Wiley and Sons.
- Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and locus coeruleus. *Journal of Vision*, 14(4), 1–20.
- Aston-Jones, G., & Cohen, J. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28, 403–450.
- Atchison, D., & Smith, G. (2000). *Optics of the human eye*. Oxford, UK: Butterworth-Heinemann.
- Beatty, J. (1982a). Phasic not tonic pupillary responses vary with auditory vigilance performance. *Psychophysiology*, 19(2), 167–172.
- Beatty, J. (1982b). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292.
- Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, 5, 371–372.
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). USA: Cambridge University Press.
- Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods*, 45, 1322–1331.
- Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. *Stereoscopic Displays and Virtual Reality Systems XI*, 5291, 36–46.
- Einhäuser, W., Koch, C., & Carter, O. L. (2010). Pupil dilation betrays the timing of decisions. *Frontiers in Human Neuroscience*, 4(18), 1–9.
- Everitt, B. S., & Dunn, G. (2001). *Applied multivariate analysis*. New York: Oxford University Press.
- Forsyth, D. A., & Ponce, J. (2011). *Computer vision: A modern approach* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods*, 43(4), 1171–1181.
- Hayes, T. R., & Petrov, A. A. (2015). *Learning is in the eye of the beholder: Phasic pupil diameter decreases during perceptual learning*. (Manuscript submitted for publication)
- Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611), 1190–1192.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, UK: Oxford University Press.
- Jay, B. S. (1962). The effective pupillary area at varying perimetric angles. *Vision Research*, 1, 418–424.

- Jennings, J. A., & Charman, W. N. (1978). Optical image quality in the peripheral retina. *American Journal of Optometry and Physiological Optics*, 55(8), 582–590.
- Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, 47(2), 310–339.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory and arithmetic tasks. *Psychophysiology*, 48(3), 323–332.
- Koss, M. (1986). Pupillary dilation as an index of central nervous system α_2 -adrenoceptor activation. *Journal of Pharmacological Methods*, 15(1), 1–19.
- Kuchinke, L., Võ, M. L., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, 65, 132–140.
- Loewenfeld, I. (1993). *The pupil: Anatomy, physiology and clinical applications*. Detroit, MI: Wayne State University Press.
- Mathur, A., Gehrmann, J., & Atchison, D. A. (2013). Pupil shape as viewed along the horizontal visual field. *Journal of Vision*, 13(6), 1–8.
- Miller, N. R., & Newman, N. J. (Eds.). (2005). *Walsh & Hoyt's clinical neuro-ophthalmology: Volume one (6th ed.)*. Philadelphia, PA: Lippincott Williams & Wilkins.
- Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, Advance online publication. doi: 10.1002/hbm.22466.
- Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index of locus coeruleus noradrenergic arousal function in humans. *Psychophysiology*, 48(11), 1531–1542.
- Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, 15(7), 1040–1046. doi: 10.1038/nn.3130
- Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59, 185–198.
- Preuschoff, K., Marius, B., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, 5(115), 1–12.
- Raisig, S., Welke, T., Hagendorf, H., & van der Meer, E. (2010). I spy with my little eye: Detection of temporal violations in event sequences and the pupillary response. *International Journal of Psychophysiology*, 76, 1–8.
- Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: Its roles in the regulation of arousal and autonomic function Part I: Principles of functional organization. *Current Neuropharmacology*, 6(3), 235–253.
- SensoMotoric Instruments. (2009). *iView X user's manual, version 2.4*. Boston, MA: Senso Motoric Instruments GmbH.
- Spring, K. H., & Stiles, W. S. (1948). Apparent shape and size of the pupil viewed obliquely. *British Journal of Ophthalmology*, 32, 347–354.
- SR Research. (2010). *Eyelink 1000 user's manual, version 1.5.2*. Mississauga, ON: SR Research Ltd.
- Tobii Technology. (2008). *Tobii X60 & X120 user manual, revision 3*. Danderyd, SE: Tobii Technology AB.
- Tobii Technology. (2011). *Tobii T60 & T120 user manual, revision 4*. Danderyd, SE: Tobii Technology AB.
- Tobii Technology AB. (2010). Tobii eye tracking: An introduction to eye tracking and Tobii Eye Trackers. (*white paper*). Retrieved from <http://http://www.tobii.com/eye-tracking-research/global/library/white-papers/tobii-eye-tracking-white-paper/>.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Võ, M. L., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology*, 45(1), 130–140.