The Receiving End of the Cooperative Eye: a test by construction

M Schmettow, C. Willemse, S. Borsci

## Introduction

The human eye region plays a central role in non-verbal communication. Beyond reading emotional states, humans are exceptionally quick and accurate at discerning one another’s gaze direction. The Cooperative Eye Hypothesis (CEH) proposes that this ability evolved alongside the emergence of human hyper-cooperation and altruism. A key anatomical feature supporting CEH is the conspicuous white sclera, which is present in Homo sapiens but absent in all other extant great apes, whose scleras are largely obscured by pigmented conjunctiva (Kobayashi & Kohshima (2001)).

In most great apes, scleral camouflage likely evolved to protect against predators by masking eye position. The CEH, however, interprets the depigmentation of the human sclera as an adaptation for signalling gaze direction, facilitating joint attention in coordinated activities such as hunting and child-rearing. Developmental studies show that even newborns respond to direct gaze and display early forms of gaze following, suggesting an innate cognitive basis for this ability (Farroni et al. (2004)).

Evolutionary theory demands that a mutation must have an immediate reproductive benefit in order to stabilize in a species. The evolution of communicative traits is often difficult to to explain because of the inherent reciprocity. A novel signal can only evolve when a matching receiver mechanism exists, and vice versa. While a specialized eye direction detector (EDD) has been proposed even before the CEH (BARON-COHEN (1995)), the conundrum to solve is which of the two emerged first.

Two extreme scenarios are possible: (1) a pre-existing gaze-reading ability in early hominids, which became more effective with scleral depigmentation, or (2) the co-evolution of the white sclera and a new cognitive mechanism for gaze perception. Comparative studies indicate that chimpanzees rely mainly on head orientation rather than eye cues, supporting the idea that human-level gaze reading is a relatively recent adaptation. A reasonable middle ground between these extremes is that early hominids may have possessed a rudimentary visual mechanism for gaze estimation, such that scleral depigmentation had some immediate benefits. Subsequent selective pressure may then have produced a specialized, more effective EDD.

We argue that mutations leading to depigmentation are more likely to happen than mutations producing novel neural pathways. Under this hypothesis, the pre-existence of a rudimentary receiver, as well as the evolution of a specialized EDD are more likely if the signal can be deciphered with general visual-cognitive functioning in humans.

The present work addresses this question in two steps. First, we test the computational simplicity of eye direction reading by introducing the QuadBright method, a eye-tracking algorithm that estimates gaze direction from visible-light brightness patterns using simple statistical learning. We evaluate its accuracy to test whether such a minimalist approach suffices for reliable gaze estimation—thereby “testing CEH by construction.”

Second, we explore the possibility that the QuadBright algorithm approximates what the human cognitive system uses for gaze perception. If so, QuadBright could serve as a candidate computational model for the hypothesized EDD. By comparing machine and human performance under identical visual constraints, we aim to test whether human gaze perception operates within these limits.

### The White Sclera in human eyes

Among all extant great apes, only Homo sapiens displays a conspicuous white sclera (Kobayashi & Kohshima (2001)). While other apes possess anatomically white scleras, theirs remain largely concealed by pigmented conjunctiva, creating minimal contrast with the iris and surrounding facial features. This distinction is particularly striking given that maintaining scleral pigmentation requires sustained metabolic investment through complex biosynthetic pathways, while depigmentation reduces energetic costs.

The common explanation for scleral pigmentation follows a camouflage hypothesis (Kobayashi & Kohshima (2001), Wolf et al. (2022)): the high contrast between a dark pupil and white sclera makes an animal more visible to predators and competitors and may even reveal the focus of attention. For species facing predation pressure or intense intraspecific competition, concealing the white sclera provides a clear survival advantage by preventing others from detecting an individual or predicting its next move. The ubiquity of pigmented scleras across great apes indicates strong, persistent selection pressure favoring camouflage despite its energetic costs.

The human pattern represents a remarkable evolutionary reversal. Despite the energetic advantage of reduced pigmentation, the conspicuous white sclera could only have evolved if it provided fitness benefits that outweighed the costs of increased visibility to predators and competitors. The depigmented sclera creates maximum contrast with the iris, making the eye region highly salient even at distance or in peripheral vision. This implies that either the original selection pressure disappeared for humans, or by another selection pressure depigmentation became so advantageous in human ancestors that it overcame the universal great ape pattern of scleral camouflage.

### The Cooperative Eye Hypothesis (CEH)

The Cooperative Eye Hypothesis proposes that scleral depigmentation evolved as an adaptation for enhanced gaze signaling, facilitating rapid joint attention in cooperative activities such as coordinated hunting, foraging (Kobayashi & Kohshima (2001)), and child-rearing (@). This hypothesis situates the unique human eye morphology within the broader context of human hyper-sociality and the emergence of unprecedented levels of cooperation among early hominids.

Supporting evidence comes from developmental studies demonstrating that human infants respond to direct gaze and exhibit rudimentary gaze-following behaviors from birth (Farroni et al. (2003), Farroni et al. (2004), Farroni et al. (2007)). These findings suggest that gaze perception operates through innate cognitive mechanisms that are functional in early development, before extensive learning could account for such abilities. The early emergence and universal nature of gaze sensitivity across human populations indicates deep evolutionary roots for this capacity.

Comparative studies with chimpanzees provide additional support for the CEH. Despite their close genetic relationship to humans, chimpanzees rely primarily on head orientation rather than eye cues when interpreting gaze direction (Tomasello et al. (2007)). This finding suggests that sophisticated eye-based gaze reading represents a derived human trait rather than an ancestral great ape characteristic, supporting the hypothesis that enhanced gaze signaling co-evolved with human cooperative behaviors.

The presence of pigmentation in chimpanzees suggests that scleral depigmentation occurred relatively recently in human evolution—likely within the past two million years following divergence from the last common ancestor with chimpanzees. However, this is not decisive, as the same is known about the EDD.

From a biological standpoint, pigmentation pathways are notably complex, involving multiple steps and enzymes—including tyrosinase and its associated proteins (e.g., OCA2, TYRP1, SLC45A2)—each essential for effective melanin synthesis and distribution (Grønskov et al., 2007). Albinism, as a phenotype resulting from mutations in any of these genes, occurs at a relatively high frequency (~1:17,000–20,000 individuals) (Bakker et al., 2022). Several non-lethal mutations are known in humans and apes Dessinioti et al. (2009), Mayhew & Gómez (2015).

Mearing et al. (2022) found that sclera pigmentation varies stronger in anthropoid ape faces than was previously held. In contrast, there is practically no variation in the human sclera

Given the high frequency of non-lethal pigmentation mutations, it is plausible that depigmentation could arise relatively easily in a population. Nevertheless, to become fixed in the population, it had to confer immediate fitness advantages, which requires a receiver mechanism.

### The Eye Direction Detector (EDD)

The existence of a specialized receiver represents a necessary corollary to the CEH. For scleral depigmentation to provide fitness benefits, human cognitive architecture must include mechanisms capable of rapidly and accurately extracting gaze direction from visual input. For BARON-COHEN (1995) eye direction detection (EDD) is a pivotal mechanism in his concept of Theory of Mind, marking it as a primarily social cognitive device.

The existence of specialized neural networks for gaze processing is supported by neuroimaging studies showing that certain brain regions are preferentially activated when individuals view eye regions or interpret gaze direction (Babinet et al. (2021)).

This raises the reciprocity conundrum: how could the white sclera produce immediate reproductive benefits? One possibility is that the white sclera encountered a rudimentary precursor of the EDD. As Laidre & Johnstone (2013) points out, it is a common assumption that “many signals have evolved from what once were cues.” Alternatively, early humans may have been able to learn to read glance directions, with the white sclera providing a signal clear enough to facilitate this learning process.

For the white sclera to become a cue, it had to had possessed visual properties that made it salient for potential receivers. Indeed, the high contrast between the white sclera and the dark pupil creates a visually striking pattern that naturally draws attention. This salience would have made it easier for early humans to detect and interpret eye movements, even in peripheral vision or at a distance.

The salience of a cue in an evolutionary context can be understood in terms of its *computational simplicity*. Under this view, the likelihood that a matching cognitive mechanism already existed depends largely on how complex this mechanism had to be. Consequently, evolutionary novel receiver mechanisms, such as the EDD, are expected to exhibit computational simplicity while maintaining functional efficiency (Guilford & Dawkins (1991)).

The simplicity requirement is underlined by the fact that gaze reading emerges in human infants before the visual system has fully developed (Babinet et al. (2021)). This suggests that the EDD relies on low-level visual cues that can be processed quickly and efficiently, rather than complex feature extraction or high-resolution image processing.

In our first study, we test the computational simplicity of the white sclera by designing a technical receiver device using minimal input and computational resources, and evaluate its accuracy compared to human gaze reading capabilities.

### Bio-plausible eye tracking

If the CEH, extended by the receiver mechanism, is correct, the human eye should be computationally simple to read: its morphology and coloration should allow for fast, accurate gaze estimation using only the visual cues available to a human observer. If the human eye is indeed a computationally simple signal, then it should be straightforward to build a technical device that can read eye directions using only visual information and simple data processing methods.

Most commercial eye-tracking devices use signals that are not available to the human visual system, such as infrared reflections or magnetic fields (Singh & Singh (2012)). The more recent video-oculographic devices instead use visible light cameras (Zhou & Geng (2004)) and recently gave rise to the development of webcam, smartphone and VR eye tracking (Ciesla & Koziol (2012)). Several algorithms have been developed in this line, often making use of advanced machine learning or computer vision techniques (Krafka et al. (2016)).

These techniques are computationally complex and demanding, especially with the partial occlusion of the pupil by the eye lids. They also require more advanced processing methods, such as filtering, edge detection or neural networks. These, in turn, are based on iterative optimization methods, which are computing intensive.

If the eye ball sends such easy to decipher signals that humans can still read glance directions in highly degraded images (Yorzinski et al. (2021)), then it must be possible to build an eye tracking device that effectively uses purely visual information by rather simple methods of data processing. While not guaranteed, it is possible that designing an eye tracking algorithm under these constraints even results in a good model, how human glance reading works.

From the characteristics of human glance reading, the following *bio-plausibility requirements* can be derived for the algorithm:

1. based on visible light
2. based on low-level visual features available in infants
3. effective in degraded images and from a distance
4. universal, requiring only minimal calibration

We introduce the *QuadBright* method to eye tracking, which employs the simple fact that the moving pupil produces a change in *horizontal and vertical brightness distribution*. The most simple approach to capture a bi-directional brightness distribution is to split the image horizontally and vertically in four quadrants (NE, SE, SW and NW, see Fig X) and taking the average brightness (Br). This effectively produces a 4 pixel grayscale image. The four quadrants are used as input for a multiple linear regression model, which predicts the horizontal and vertical position of the eye ball. The following equation shows the model for the horizontal eye position, the vertical position is analog.

### YET Zero prototype

To test the QuadBright algorithm, a simple eye tracking device was constructed, called Yet Zero (YET = Yet is your Eye Tracker). The main design goal was to create a low-cost, easy-to-build device that can be used in various experimental settings.

Most important for the physical design of the eye tracking device is a small footprint of the camera to minimize obstruction of the field of view when it is mounted in frontal position. In contrast, the resolution of the camera barely matters, as QuadBright essentially compresses the input into a four pixel frame. The choice fell on a commercially available USB endoscope camera with a native resolution of 640x480 at 24Hz refresh rate and an AppoTech AX2311 video controller. With a diameter of 5.5mm this camera creates very little obstruction when mounted in the visual field and the six dimmable LED lights provide a stable light source for better accuracy. A simple 3D-printed socket was created to be able to glue the camera to a stick, which in turn was connected to an improvised head mount, using a headphone ([Figure 1](#fig-yet-zero)).

The YET Zero application was written in Python, using PyGame for the GUI and basic image processing routines from OpenCV (Bradski, 2000). The regression model was implemented using the Scikit-learn library (Pedregosa et al., 2011). As usual in eye tracking, a calibration phase trains the model on Quadbright data obtained on (nine) initial calibration points. To account for small deviations between trials, single-point quick calibrations are performed to realign the parameters and .

In first tests this system seemed worked well, but it was susceptible to changes in computer screen brightness introduced by the presented stimuli. While the best solution is to use a frontal brightness sensor and extend the model accordingly, as a quick fix the quick calibration was augmented with a highly degraded preview of the next stimulus, just enough to keep the brightness level stable.

|  |  |
| --- | --- |
| |  | | --- | | (a) Headmount with camera | |

Figure 1:  YET Zero prototype assembled from a USB endoscope camera, a ruler, a headphone and a kitchen roll as headrest (Bender & Schmettow, 2024).

# Study 1: Accuracy of the QuadBright algorithm

The first study was conducted to evaluate the accuracy of the QuadBright algorithm. The algorithm was tested on a 36 participants, who were asked to glance at an orange target on a white background, moving in a random pattern between 24 positions in a six-by-four grid.

## Methods

The experiment was carried out with the YET Zero prototype. A headmount was constructed from a headphone and a kitchen paper roll was used as a chin rest. Participants were seated in 45cm viewing distance using a 16:9 screen with a 41cm diagonal. The experiment consisted of 24 trials per condition. The target moved in a random pattern between 24 positions on the screen. The participants were asked to follow the target with their eyes. The QuadBright algorithm tracked their eye movements and the accuracy of the algorithm was measured by calculating the angular standard error of predicted eye movements of the participants.

The original experiment tested three lighting conditions (visible, infrared, no light), as well as consistent/inconsistent screen brightness levels at quick calibration. As expected, inconsistent brightness levels led to a strong decline in accuracy, whereas the light source conditions differed very little. For the analysis here, data from inconsistent background brightness was discarded and the lighting conditions were pooled. Full details of the experiment can be obtained from Bender & Schmettow (2024).

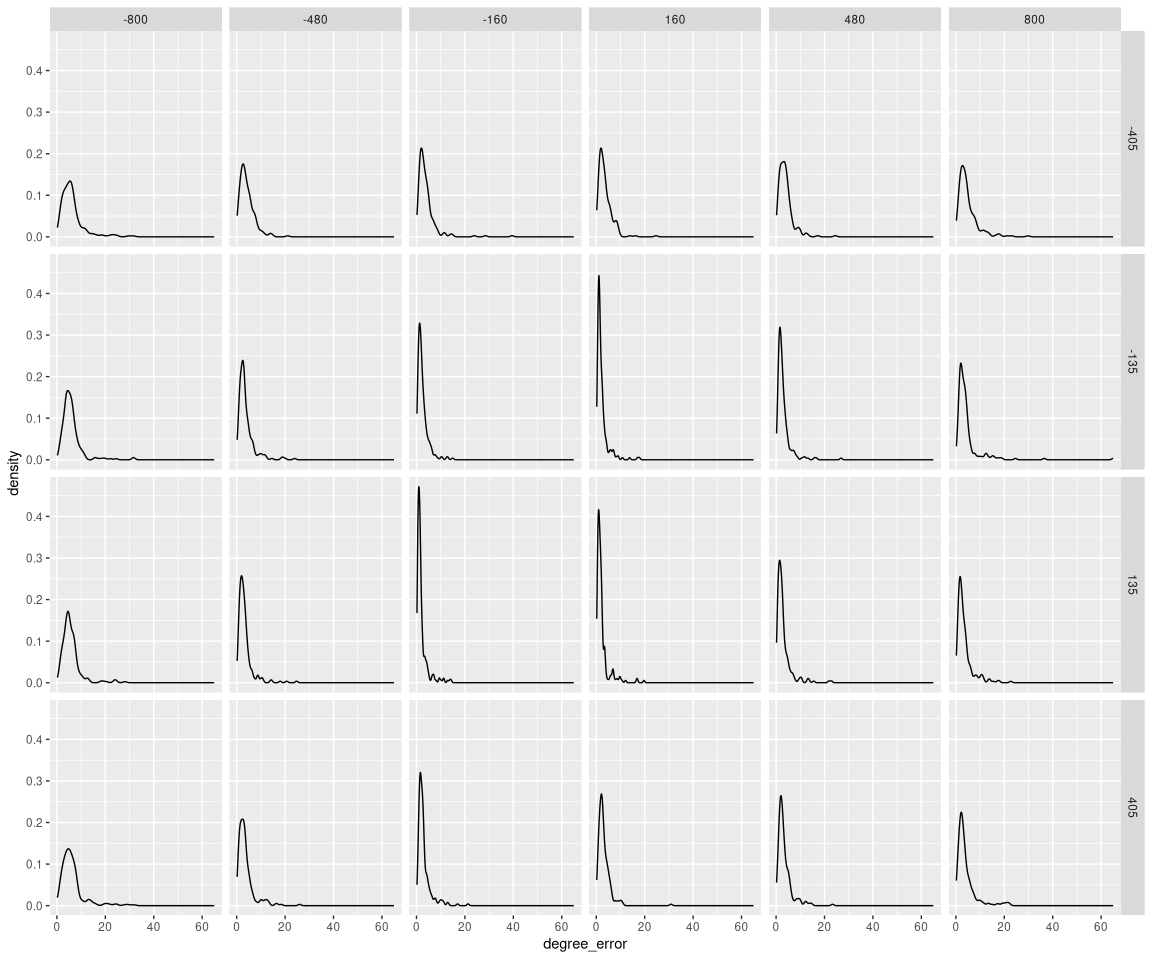
A multi-level linear model was used to analyze the data, with random intercepts for participants and target positions. The dependent variable was the degree error, calculated as the angular distance between the actual target position and the predicted gaze position. The model was fitted using the lme4 package in R.

|  |
| --- |
| Figure 2: Measurement error by target position. Histograms show the distributions of angular errors (in pixel), with a density distribution shown in Red. Blue vertical lines indicate the median accuracy. Note the logarithmic scale on the x-axis. |

## Results

The effectiveness of the Quadbright method is measured as degree errors between real target position and measured position. The sample median of degree errors 2.7 degree, with a mean of 3.2 degree. 95 percent of all measurements are below 10.3 degree error. Extreme errors mostly occurred on three (out of 36) participants.

|  |
| --- |
| Figure 3: Distribution of degree errors across all trials and participants. Vertical lines indicate the mean (green), median (blue) and 95% quantile (red). Note the logarithmic scale on the x-axis. |



We ran a polynomial multi-level model with Intercept to map angular errors to target positions. To account for the curvature of measurement error across the screen, we used a second-order polynomial term, where the intercept represents the point of minimum angular error in screen center. To account for individual differences, we included the same term on participant level (random effects).

In the center of the screen the error is xy. Both square terms are positive, which means the error increases towards extreme positions. However, the linear terms are both positive, which means the point of minimum error is xy to the right and xy up from screen center, with quite some uncertainty.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Coefficient estimates with 95% credibility limits   | fixef | center | lower | upper | | --- | --- | --- | --- | | Intercept | -1.2614164 | -2.7666335 | 0.4749971 | | x\_linear | -0.0009461 | -0.0012579 | -0.0006482 | | x\_square | 0.0000036 | 0.0000025 | 0.0000050 | | y\_linear | -0.0006679 | -0.0011998 | -0.0001857 | | y\_square | 0.0000049 | 0.0000029 | 0.0000069 | |

The model predicts the degree error in the center of the screen to be 2.3 degree, with a high level of certainty ([Table 1](#tbl-fixef-M1)). The square terms for both axis are positive, which means that the error increases towards extreme positions. The linear terms are both negative, which means there is a lateral difference, with the largest error in the lower, mount-side (non-nasal) corner.

[Figure 4](#fig-predicted-degree-error) shows interpolated degree errors per participant. While for all participants the horizontal and vertical errors is lowest near the center region, there appear to be strong individual differences in the curvature. At the same time, there is a surpising lack of crossings between the individual curves, which means that participants with low error in the center region also have low error in extreme positions.

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | (a) Horizontal | | |  | | --- | | (b) Vertical | |

Figure 4: Predicted degree error by target position and participant. Points show center estimate per trial with LOESS lines for interpolation.

In conclusion, eye tracking using the Quadbright algorithm typically produces accuracy in the range of 2-5 degree error. This is a magnitude above what is achieved by commercial eye tracking devices using corneal reflections (e.g. Pastel et al. (2021)), but in the same range as other video-oculographic devices (Heck et al. (2023)). In the study of Gale & Monk (2000) similar levels of accuracy were observed on participants when they judged glance directions from eye and head positions. Another similarity is the effect of eccentricity on accuracy, which has been observed for human participants by Loomis et al. (2008). Altogether, the Quadbright device qualifies as a candidate model for human eye direction reading.

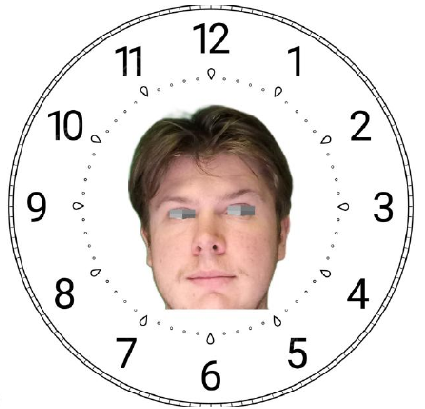
## Study 2: Human glance perception with QuadBright information

The second study tested the possibility that human direction reading uses a Quadbright device by a gaze reading experiment, where accuracy was measured in a control condition and in a condition where eye regions were reduced to four brightness levels. To approximate the depth of processing in both conditions, exposure times were varied from 70 milliseconds to one second.

### Experimental design

In this experiment, participants see a frontal face glancing at the hours positions on a clock face (**?@fig-clockface**) and are asked to read the correct hour position. The Quadbright condition was created by masking the eyeball region by a two-by-two grid of brightness levels (**?@fig-X-b**). A Python script using the OpenCV library (Bradski, 2000) was used to automate the task (Großerichter & Schmettow, 2025). Exposure times varied in five steps (70, 140, 400, 600 and 1000 milliseconds).

|  |  |
| --- | --- |
| |  | | --- | | (a) Control condition | |

Figure 5:  Example stimuli image with eye direction at 3 o’clock position (Hölter & Schmettow, 2025).

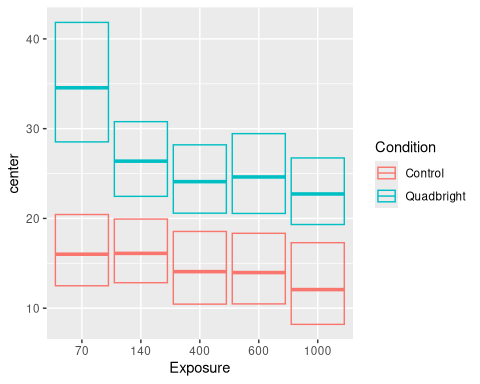
### Sample

### Data analysis

The outcome variable is the deviations between true and reported hour positions, resulting in a discrete outcome. A two-factorial linear term with conditional effects for exposure time and QuadBright degradation was used for population-level (fixed effects), as well as on Stimulus and Participant level (random effects). To account for over-dispersion, a negative-binomial outcome distribution with logarithm link function was tried first, but the extremely large reciprocal dispersion parameter suggested that the data is not over-dispersed and the final model was therefore fitted using the Poisson family. All computations were carried out using the Brms package (Bürkner, 2018, version 2.22.0) package in R (R Core Team, 2025, version 4.5.1).

### Results

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2   | Population-level effect | center | 2.5% CrI | 97.5% CrI | | --- | --- | --- | --- | | Intercept | 15.8° | 11.9° | 20.9° | | Quadbright | 222% | 170.7% | 288.7% | | 140 | 99.7% | 89.4% | 111.5% | | 400 | 90.6% | 81.6% | 100.5% | | 600 | 89.8% | 80.3% | 100.1% | | 1000 | 81.9% | 69.9% | 93.3% | | Quadbright:140 | 75.3% | 65.6% | 86.2% | | Quadbright:400 | 75.7% | 65.7% | 87.2% | | Quadbright:600 | 78.4% | 67.3% | 91.4% | | Quadbright:1000 | 78.6% | 67.6% | 93.1% | |



In [Table 2](#tbl-fixef-M4) the Intercept is the average error in the baseline condition (Control, 70 milliseconds), which is around 16° with moderate certainty. In the control condition, error rates remain stable (99.7%) with 140 ms exposure, but with 400ms and 600ms the error rate drops to around 90% and finally to 81.9%.

In the Quadbright condition with 70 ms exposure, the error is around twice as high compared to the control condition, although with considerable uncertainty. Initially, the expected degree error is around 35°. Different to the control condition, error rates take the first drop at 140ms to around 75%. From that point on, the error rates drop similarly to the control condition.

If human eye direction reading is similar to a Quadbright device, we expect that it is a universal trait in the population. Figure [Figure 6](#fig-participant-trajectories) shows the participant-level performance trajectories. Initially, all of them showed lower performance in the Quadbright condition. The second dominant effect that can be observed is an enormous variance in the Quadbright conditions. Initially, the sample appears to fall into two clusters, one in the region below 40 degree errors, and a few participants dramatically failing with Quadbright signals. Starting with 140ms, a third group appears with almost stable performance around 10 degree.

|  |
| --- |
| Figure 6 |

A related question is whether the population-level pattern is representative for all displayed eye directions on stimulus level. [Figure 7](#fig-stim-trajectories) shows that this is not so.

|  |
| --- |
| Figure 7 |

## Discussion

### Evolution of the effective sender

The Cooperative Eye Theory implies that the eyeball is a very effective sender of glance direction signals, as simple receiver mechanisms are more likely to evolve. By construction of the most simple eye tracking device that can come to mind, we tried to establish a new limit on how effective the sender is.

Overall, the Quadbright device showed very similar characteristics as human observers. Accuracy is in the same range as observed by Gale & Monk (2000), and the effect of eccentricity is also similar. In comparison with any other method of eye tracking the Quadbright device is extremely simple, with a 2x2 array of brightness sensors as input requirements.

Quadbright is extremely efficient in memory consumption, as it works with four brightness levels, which makes a total of 4 byte in the input buffer. Computationally, Quadbright is extremely efficient, as it uses a linear model with only 12 parameters to calculate eye directions. By the method of minimizing the sum of squares, all calculations in the process have closed solutions, requiring no costly numerical approximations, recursions, or iterative procedures. From a biological perspective, every organism that can sense horizontal and vertical brightness changes, can evolve this function, or acquire it by learning.

We may not forget here, that the spherical form of eye balls has evolved not for computational reasons, but for the physical requirements of creating a moving, variable-focus optical lens from a sack of pressurized liquid. The second salient cue of eyeballs is the dark iris, which adjusts the opening of the lens under varying light levels. Pigmentation in the iris seems to reduce stray light effects, where internal reflections reduce contrast levels (Kruijt et al., 2011).

### Evolution of the receving end

The simplicity of the Quadbright device supports the idea of sclera camouflage as a default condition among apes, despite the metabolic effort. It is hardly conceivable that apex predators, and in particular big cats, would not evolve the ability to read the eyes of their prey. This ability is not necessarily eye direction reading as observed in humans. Another scenario is that first evolved the ability to detect faces. Animals equipped with face detection and sufficiently large brains are likely to *learn by experience* that brightness changes in the eye region often precedes an escape.

Is the Quadbright algorithm equivalent with human glance reading? Is input organized as an image of four areas and the visual system performs a series of simple linear equations? If Quadbright is a good model of human glance perception, then all humans should be able to read glance directions from brightness four grey areas. Our second study showed that Quadbright degradation led to a strong overall decline in reading performance. Quadbright is definitely not the mechanism that drives modern human eye direction reading.

While most participants were less accurate with degraded images, the majority of the participants was still much beyond guessing. Crucially, the Quadbright degradation makes it almost impossible for any mechanism that goes beyond brightness gradients, such as edge detection. We can conclude that a minimum of two processing modes for eye direction reading exists. The first one is based on brightness gradients with moderate accuracy, and the second one extracts higher-level features for high accuracy.

A multiple receiver model has been proposed by Anderson et al. (2016), who argue that there may be different mechanisms for precise gaze triangulation and coarse judgement of direction. By using a contrast reversal manipulation they showed that the first depends on geometric cues, whereas the second depends on luminance and motion cues.

The multiple receiver model is also inline with the evolutionary process where a rudimentary cue receiver paves the way for the co-evolution of signals and more advanced receivers. Brightness gradient processing is not only a good candidate for such a proto-receiver, and there is no reason why this ability should vanish, especially when it still fulfills functions that the more advanced receiver cannot, such as long distance reading and peripheral change detection.

Interestingly, some participants were not affected by the Quadbright image reduction and by tendency even performed better on a good level. This has two possible explanations: One is that Quadbright processing can be learned, and these individuals somehow have acquired this skill prior to the experiment. Under the multiple receiver hypothesis one can also speculate that these individuals do not possess the foveal receiver, but learned to compensate almost fully using brightness processing.

When BARON-COHEN (1995) postulated an EDD, he argued that gaze reading is not just a utilitarian function for shared attention, but a primary cue for Theory-of-Mind (TM), i.e. the ability to attribute mental states to others. In (**Cohen\_autism\_spectrum?**) he proposes a connection between autism spectrum disorders and a dysfunctional TM. If the lack of an EDD is part of this dysfunction, then we would expect diagnosed participant to show equal performance in the conditions of our experiment.

## Limitations

A limitation of the first study is that the YET Zero prototype fixes the camera close to eye, which does not match the conditions of a human observer. However, extreme data reduction of Quadbright input suggests that the algorithm will work for distant images as well. While the Quadbright device is not robust to adverse lighting conditions, this may not limit the conclusions of the study, because humans light adjustment capabilities of the human visual system (and that of many mammalian clades) are superb. A straight-forward technical approximation could be the addition of a few simple brightness sensors, and add these signals as control variables to the linear model. Finally, face detection and object tracking are required to find the eye region in a face. However, these are well-established techniques in computer vision, and the human visual system is extremely good at these tasks.

### Future research

Speculating about how receiver mechanisms evolved, and how many there exist is fascinating, but much more research needs to be conducted to make good for the fact that you cannot truly experiment with evolution. The eye tracking device, we developed, can be used to emulate other algorithms, for example using more fine-grained brightness distributions (“Hexabright”) or processing geometric cues.

The dominant effect of the Quadbright manipulation is how differently it acts on individuals. The divide spans two magnitudes and must therefore have correlates in everyday performance. We would expect that performance in the Quadbright condition is correlated with the ability to read glances under poor conditions, for example with peripheral viewing or when the sender is wearing glasses.

Studying how eye direction cues establish shared attention also has impact in at least two modern fields of application. One open problem in transport is how automated vehicles can effectively communicate attention and intentions with human traffic members. For example, making eye contact is often observed in communication between drivers and street-crossing pedestrians. External human-machine interfaces (e-HMI) have been tested for communication on the outside of cars, using text, icons and facial expressions. Understanding the role and processing modes of eye direction reading can be used to design more efficient signals.

Another field of application is the design of artificial faces. In social robotics, robot faces are often designed to be expressive, and the role of glance reading in human-robot interaction is well established. However, designing more human-like eye regions for this purpose may result in an adverse effect known as the Uncanny Valley. Artificial faces that are too human-like, but not quite right, are perceived as creepy. The results of this study suggest that degraded eye regions can convey glance directions, which could facilitate the quality of interactiuon, without falling into the Uncanny Valley.

Finally, the eye tracker we build to test our ideas about the receiving end has already been used extensively for bachelor-level student projects. With some refinement it could deliver sufficient accuracy for many psychological experiments and applied research studies, with the added benefit of being much easier to replicate for everyone than with using commercial eye tracking devices.

# References

Anderson, N. C., Risko, E. F., & Kingstone, A. (2016). Motion influences gaze direction discrimination and disambiguates contradictory luminance cues. *Psychonomic Bulletin and Review*, *23*, 817–823. <https://doi.org/10.3758/s13423-015-0971-8>

Babinet, M.-N., Cublier, M., Demily, C., & Michael, G. A. (2021). Eye direction detection and perception as premises of a social brain: A narrative review of behavioral and neural data. *Cognitive, Affective, &Amp; Behavioral Neuroscience*, *22*(1), 1–20. <https://doi.org/10.3758/s13415-021-00959-w>

Bakker, R., Wagstaff, E. L., Kruijt, C. C., Emri, E., Karnebeek, C. D. M. van, Hoffmann, M. B., Brooks, B. P., Boon, C. J. F., Montoliu, L., Genderen, M. M. van, & Bergen, A. A. (2022). The retinal pigmentation pathway in human albinism: Not so black and white. In *Progress in Retinal and Eye Research* (Vol. 91). Elsevier Ltd. <https://doi.org/10.1016/j.preteyeres.2022.101091>

BARON-COHEN, S. (1995). The eye detection detector (EDD) and the shared attention mechanism (SAM) : Two cases for evolutionary psychology. *Joint Attention : Its Origin and Role in Development*, 41–59. <https://cir.nii.ac.jp/crid/1572261550899222144>

Bender, F. R., & Schmettow, M. (2024). *Evaluaton of the low budget do it yourself eye tracking system: Your eye tracker (YET) human factors and engineering psychology* [PhD thesis, University of Twente]. <https://purl.utwente.nl/essays/104538>

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb’s Journal of Software Tools*.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>

Ciesla, M., & Koziol, P. (2012). Eye pupil location using webcam. *Computer Vision and Pattern Recognition*, 1–11.

Dessinioti, C., Stratigos, A. J., Rigopoulos, D., & Katsambas, A. D. (2009). A review of genetic disorders of hypopigmentation: Lessons learned from the biology of melanocytes. *Experimental Dermatology*, *18*(9), 741–749. <https://doi.org/10.1111/j.1600-0625.2009.00896.x>

Farroni, T., Mansfield, E. M., Lai, C., & Johnson, M. H. (2003). Infants perceiving and acting on the eyes: Tests of an evolutionary hypothesis. *Journal of Experimental Child Psychology*, *85*(3), 199–212. <https://doi.org/10.1016/s0022-0965(03)00022-5>

Farroni, T., Massaccesi, S., Menon, E., & Johnson, M. H. (2007). Direct gaze modulates face recognition in young infants. *Cognition*, *102*(3), 396–404. <https://doi.org/10.1016/j.cognition.2006.01.007>

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy*, *5*(1), 39–60. <https://doi.org/10.1207/s15327078in0501_2>

Gale, C., & Monk, A. F. (2000). Where am i looking? The accuracy of video-mediated gaze awareness. *Perception & Psychophysics*, *62*, 586–595. <https://doi.org/10.3758/BF03212110>

Grønskov, K., Ek, J., & Brondum-Nielsen, K. (2007). Oculocutaneous albinism. *Orphanet Journal of Rare Diseases*, *2*. <https://doi.org/10.1186/1750-1172-2-43>

Großerichter, J., & Schmettow, M. (2025). *Human gaze estimation : Comparing accuracy in naturalistic and quad brightness conditions* [Thesis, University of Twente].[https://purl.utwente.nl/essays/106716]( https://purl.utwente.nl/essays/106716)

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, *42*, 1–14. <https://doi.org/10.1016/S0003-3472(05)80600-1>

Heck, M., Becker, C., & Deutscher, V. (2023). Webcam eye tracking for desktop and mobile devices: A systematic review. *Proceedings of the 56th Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2023.825>

Hölter, J., & Schmettow, M. (2025). *How do humans perceive gaze direction* [Thesis, University of Twente].[https://purl.utwente.nl/essays/107106]( https://purl.utwente.nl/essays/107106)

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, *40*(5), 419–435. <https://doi.org/10.1006/jhev.2001.0468>

Krafka, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., & Torralba, A. (2016). Eye tracking for everyone. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <http://gazecapture.csail.mit.edu.>

Kruijt, B., Franssen, L., Prick, L. J. J. M., Vliet, J. M. J. van, & Berg, T. J. T. P. van den. (2011). Ocular straylight in albinism. *Optometry and Vision Science*, *88*, E585–E592. <https://doi.org/10.1097/OPX.0b013e318212071e>

Laidre, M. E., & Johnstone, R. A. (2013). Animal signals. *Current Biology*, *23*, R829–R833. <https://doi.org/10.1016/j.cub.2013.07.070>

Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, *37*(9), 1443–1457. <https://doi.org/10.1068/p5896>

Mayhew, J. A., & Gómez, J. (2015). Gorillas with white sclera: A naturally occurring variation in a morphological trait linked to social cognitive functions. *American Journal of Primatology*, *77*(8), 869–877. <https://doi.org/10.1002/ajp.22411>

Mearing, A. S., Burkart, J. M., Dunn, J., Street, S. E., & Koops, K. (2022). The evolutionary drivers of primate scleral coloration. *Scientific Reports*, *12*(1). <https://doi.org/10.1038/s41598-022-18275-9>

Pastel, S., Chen, C. H., Martin, L., Naujoks, M., Petri, K., & Witte, K. (2021). Comparison of gaze accuracy and precision in real-world and virtual reality. *Virtual Reality*, *25*, 175–189. <https://doi.org/10.1007/s10055-020-00449-3>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Singh, H., & Singh, J. (2012). Human eye tracking and related issues: A review. *International Journal of Scientific and Research Publications*, *2*. [www.ijsrp.org](https://www.ijsrp.org)

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, *52*, 314–320. <https://doi.org/10.1016/j.jhevol.2006.10.001>

Wolf, W., Thielhelm, J., & Tomasello, M. (2022). Five-year-old children show cooperative preferences for faces with white sclera. *Journal of Experimental Child Psychology*, *225*, 105532. <https://doi.org/10.1016/j.jecp.2022.105532>

Yorzinski, J. L., Thorstenson, C. A., & Nguyen, T. P. (2021). Sclera and iris color interact to influence gaze perception. *Frontiers in Psychology*, *12*. <https://doi.org/10.3389/fpsyg.2021.632616>

Zhou, Z.-H., & Geng, X. (2004). Projection functions for eye detection. *Pattern Recognition*, *37*, 1049–1056. <https://doi.org/10.1016/j.patcog.2003.09.006>