Perception of Motion

Snehesh Shrestha
snehesh@umd.edu

Yehuda Katz
yakatz@cs.umd.edu

Virinchi Srinivas
virinchi@cs.umd.edu

Matt Goldberg
mdgold@cs.umd.edu

Michelle Mazurek
mmazurek@cs.umd.edu

Cornelia Fermüller
fer@cfar.umd.edu

ABSTRACT

The goal of this study is to systematically test the parameters that create the movement in the illusion "The Enigma" and verify that perception of the illusion with these parameters is universal. The famous illusion "The Enigma" was painted by Isia Léviant in 1981. When staring at the center, the concentric circles appear to have rotatory motion even though the image is completely static. Fermüller et. al. provided a mathematical model explaining the reason the illusion is visible. By measuring how long it takes people to see movement when they look at images with different densities and line angles, we show that while the density plays an important role, the angle of the lines does not seem to influence the illusion as was previously believed. We show that uncorrected optical defects play a role in the ability and the speed at which the illusion is seen and we show that while gender and ethnicity do not to contribute towards this phenomenon, participants over the age of 60 appear to show a sudden loss of perception of illusion.

KEYWORDS

optical illusion, perception, vision, motion, human

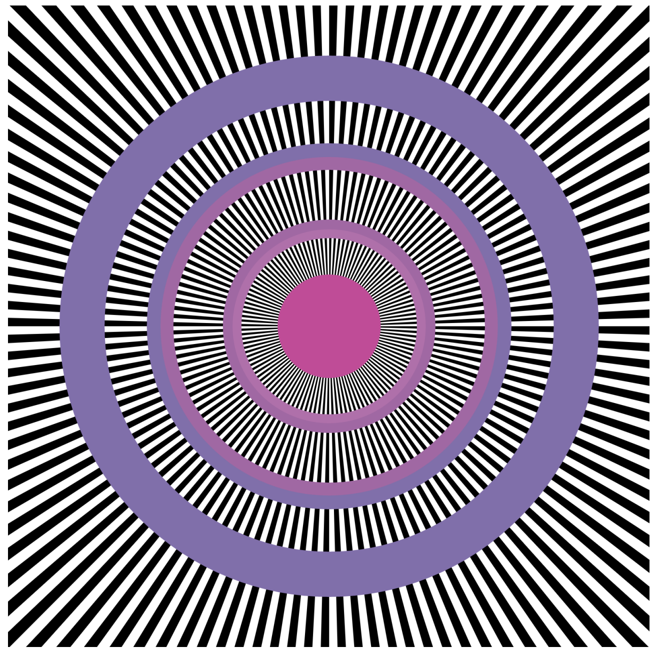

Figure 1: The Enigma by Leviant

1 INTRODUCTION

For almost four decades, people have been curious about the cause of the illusion "The Enigma" [4, 5], a stationary pattern producing illusory movement. This static image consists of black and white radial lines with three concentric purple circles superimposed on top of the lines. People viewing the illusion generally see rotatory motion along the concentric circles. Some see rotary motion in alternating directions while others see flickering or unidirectional motion. This is particularly interesting because neither the image nor the observer's head needs to be moving to see the illusion.

2 RELATED WORK

Previous research provides two opposing explanations for this illusion. Zeki et. al. [7] believe this illusion happens in the brain, as viewing the illusion activates the same brain area as when one perceives real motion. Troncoso, et. al. [6] believe the illusion is created in the retina due to tiny involuntary eye movements called microsaccades. Fermüller et. al. [2, 3] take a mixed approach where they provide a computational model with mathematical explanation of why these kinds of illusions are perceived. This same model is responsible for how humans perceive three-dimensional motion. Fermüller [3] presents a family of static images where similar illusions were created to demonstrate the validity of their theory. Beyond the researchers themselves, no human-subjects research was conducted to confirm that this phenomenon is universal. We wanted to investigate if perception of this optical illusion varies in individuals and how the perception varies as we change the patterns away from those known to be part of the original mathematical model.

Our work focuses on a human-subjects study to validate Fermüller's [3] theory that this phenomenon is universal. We further explore the role of parameters such as density of lines, angle of intersection, and demographics in the degree of appearance of the illusion.

3 METHODOLOGY

Past research on the class of optical illusions similar to The Enigma has been conducted on very small groups of people. Because our goal was to see previous findings in this area are universal, it was important to accurately conduct our experiment on a larger representative sample. For participants who observed this illusion, we focused on the effects of variations in the image and demographics that might contribute further to the phenomenon. We use elapsed time from starting to look at the image until seeing the illusion as a direct correlation of difficulty of seeing movement in the static images.

We referred to Fermüller [3] and Bach [1] to develop our MATLAB code that generated the illusion and all the other variations.

3.1 Data Collection

Because our stated goal was to reach as large and diverse a sample as possible to validate (or disprove) the parameters of seeing this illusion, we conducted an IRB-approved online study. Participants were shown various generated images, some which we expected to have optical illusions and some which did not. We measured how long it took for a participant to start seeing an illusion and some information about what illusion they saw.

3.1.1 Recruitment. We recruited on social media (Facebook, Twitter, and LinkedIn) and other person-to-group communication methods (email lists). We asked participants to do the survey only once, but we did not do any type of duplicate detection. We advertised that participants needed to be at least 18 years old and use a desktop computer or large tablet with a keyboard, and it was not possible to complete the reaction-time test — and therefore proceed with the study — without a keyboard, but we did not create any other check to enforce those requirements. Participants were not paid.

The final page of the study had sharing links for Facebook, Twitter and other social media websites and participants were asked to share the study with their friends. We embedded Open Graph metadata so that social media sites would show a teaser image (Fig. 9) and the clickbait-style subtitle of the project.

More than half of visitors to the study website had no referrer information. Of the visitors whose browsers supplied referrer information, almost half came from Facebook's mobile website, showing they were browsing on a smartphone and were therefore unable to participate in the study. The second most common referrer was Facebook's standard website. We could see using Facebook publisher tools that our link was shared on Facebook 38 times. We received a handful of visits from LinkedIn users, many of them also using LinkedIn's mobile app. We did not receive any visits with a Twitter referrer.

We planned to detect mobile users based on their small screen size and on the first page of the study show a message asking the user if they wanted an email reminder to complete the study at a computer, but we did not implement that in the end.

3.1.2 Study Procedure. Users who came to the website were presented with instructions telling them they needed to be using a computer with a real keyboard[1] and to sit about an arm's length from their monitor. They were shown and had to agree to a standard consent form. We told participants that we might ask for contact information in case we wanted to follow up with them, but we did not collect that information in the end. Javascript was required to complete the study.

To get a baseline for participants' reaction time to the actual optical illusions and to rule out variations among different computers, we first asked participants to complete a reaction time test. The test told participants to put their finger on the letter 'J' key on the keyboard and press to start. They were told when the word "PRESS" appeared on the screen (after a short, but random delay) to press the letter 'J' again. The reaction time test collected five samples. If the participant pressed 'J' before the word 'PRESS' appeared, they were asked to start that particular test again.

Participants were then shown 22 images with and without illusions in a randomized order. For each image, as soon as the entire webpage had loaded (to eliminate variation caused by slow network connections), participants were asked to press 'J' to start. Participants were asked to press 'J' as soon as they saw something moving in the image or if they felt they had waited long enough and not seen anything moving. When they pressed 'J', a menu would show up asking them why they pressed 'J' with the choices: 1) Did Not See Any Movement, 2) Saw Spinning Motion, 3) Saw Flickering Motion, 4) Pressed by Accident, and 5) Saw Something Else.

After two minutes with no input, the menu would show up along with a message that said "You have been looking at this image for a while? What do you see?". If we detected that the participant changed tabs or minimized the browser, we would show an overlay on the page asking the participant to refresh the page and start the test on the particular image from the beginning. Participant session data was stored in an encrypted cookie, so refreshing the page would not change the image the participant would be shown.

Participants were asked for some demographic information, including age, race, whether they need corrective lenses, whether they were wearing corrective lenses, and what kind of correction the lenses were for. They were then prompted to share the study with friends on social media. We created a click bait-style subtitle for the project, but it is not clear if it helped (or hurt) participant recruiting.

We used Piwik, an open-source, self-hosted Google Analytics clone, to collect additional visitor information including screen resolution, browser and OS versions, IP address, referrer, and time of visit in local timezone. We integrated the analytics script with the study code so we could easily match the data in the two databases.

3.1.3 Privacy. The data we collected is not particularly sensitive even if it could be traced back to a single person. Additionally, social networks are full of reaction-time-testing games which have no oversight of any kind and yet are very popular, so we believe participating in a study like this would be unlikely to cause a participant who uses social media to give us any more information than they would give to a stranger.

3.1.4 Collection Period. We opened the survey to participants on May 7, 2017. We took a snapshot for analysis on May 10, 2017, but have not closed it to new participants. We wanted to have at least 400 participants, but we did not make this goal in the short time the study was running.

3.2 Analysis

3.2.1 Metrics. We performed a quantitative data analysis. We used 4 important metrics for experiment : (a) line density (LDN) - number of lines passing through the center, (b) angles (ANG) - angle at which lines intersect concentric circles, (c) line space ratio (LDS) - ratio of lines and spaces between them, (d) reaction time (RT) - time taken for the participant to (not) observe any illusion from the time they first saw the image. Further, we also subtracted

[1] **Note: Keyboard vs. Mouse Input** A significant amount of data we collected was concerned with the time it took for the participant to see something and react. So it was important to collect timing information in a consistent way. Timing mouse clicks poses a number of challenges, for example the time taken to move the mouse pointer. When capturing the keyboard input, there is no positioning necessary and therefore keyboard input is more accurate.

the average of base reaction time (BRT) (average of best 3 times out if 5 [2]) from the reaction time for individual images. The RT we will refer to has the BRT subtracted from it. [3] stated that an illusion is observed when the angle at which lines intersect the circles is $\approx 90°$. Further, they also observed that line density plays a prominent role in the possibility of observing an illusion. Hence, we decided to use the above metrics and in particular focus on understanding how varying any metric affects the RT. Some of the metrics we used were demographics related namely age, gender, race and optical defect. All our metrics were categorical, except RT which was continuous.

3.2.2 Hypothesis Tests. Given the metrics, we framed the following research questions:

(1) Does variation in ANG affect the RT to observe any illusion?
(2) Does variation in LDN affect the RT to observe any illusion?
(3) Does variation in LDS affect the RT to observe any illusion?
(4) Do Age, Race, Gender and optical defect affect the possibility of observing an illusion?

For every research question that we framed, the corresponding null hypothesis was as follows:

(1) **H1:** Variation in ANG does not affect the RT to observe any illusion.
(2) **H2:** Variation in LDN does not affect the RT to observe any illusion
(3) **H3:** Variation in LDS does not affect the RT to observe any illusion
(4) **H4:** Age, Race, Gender and optical defect are not related to the possibility of observing an illusion (number of images with illusion seen).

For the case of hypotheses **H1**, **H2** and **H3**, we observed that each IV (LDN, LDS and ANG) was categorical and that the DV (RT) was a continuous numeric variable. To assess whether a parametric or nonparametric test would be appropriate, the Shapiro Wilk normality test was performed on each independent variable. The results of these tests yielded statistically significant p-values — in all cases below 0.05, and in the majority of the cases below 1e-5 — which ruled out the possibility of using a parametric test. Given the within subject design of the experiment and varying image parameters, we identified the non-parametric Friedman's ANOVA test as a suitable test. The assumptions of this test were satisfied by the criteria discussed above.
For the vase of hypothesis **H4**, we observed that all the IVs were categorical and DV was numeric. We used a linear regression model to perform this test, assuming linearity between each IV and DV and also a normal distribution of the residuals.

3.2.3 A Priori Power Analysis. We performed a power analysis prior to any data collection, after deciding on the four hypotheses we discussed earlier. The purpose of an a priori power analysis was to estimate the number of valid samples that each hypothesis test would require to have a at least 80% power, significance level of 0.1 and medium effect size. For each of the hypotheses, we described the number of samples as follows:

(1) LDN variation vs RT: LDN is a categorical IV with 6 levels and RT is a numeric DV. The number of degrees of freedom for performing a Friedman's ANOVA test for this hypothesis test is $(6 - 1) = 5$. This test requires a sample size of at least 117.
(2) ANG variation vs RT: ANG is a categorical IV with 5 levels and RT is a numeric DV. The number of degrees of freedom for performing a Friedman's ANOVA test for this hypothesis test is $(5 - 1) = 4$. This test requires a sample size of at least 108.
(3) LDS variation vs RT: LDS is a categorical IV with 4 levels and RT is a numeric DV. The number of degrees of freedom for performing a Friedman's ANOVA test for this hypothesis test is $(4 - 1) = 3$. This test requires a sample size of at least 98.
(4) Demographics vs Number of Illusion Perceived: Demographics include 4 IV's namely Age, Race Gender, and Optical defect with 6, 6, 3 and 12 degrees of freedom respectively. Hence, our model had a total of 24 degrees of freedom in total needing 143 samples for the prior set power.

Hence, a priori power analysis showed that we would need at least 98 samples for at least 80% power, significance level of 0.1 and medium effect size for conducting the hypothesis tests.

3.3 Limitations

Some of the limitations of our experiment are as follows:

- Sample Size : A priori power analysis shows that we need at least 97 samples for at least 80% power, significance level of 0.1 and medium effect size. However, we have 67 samples.
- Representative Sample Distribution : We could not obtain an uniformly distributed sample across all demographic parameters.
- Screen : We were not able to control screen size, lighting and brightness, and distance from screen; we assumed that they did not affect the reaction time.
- Base reaction time could not be monitored, as we did not perform the test when we were physically around.
- We could not control each participant's environment.
- There was no way for us to validate the universality claim.
- No interaction between IVs : Demographics such as age, gender, region, race, optical deficiencies etc. do not have effect on each other
- We could not verify if a participant answered questions truthfully.

4 RESULTS

4.1 Demographics

Out of the 260 participants, only 67 were valid participants who completed the test. Of the 67 participants, there were no duplicates based on IP and screen resolution. The distribution of the participants whose data we have analyzed are as follows: Participants were observed to be roughly evenly split by gender. Age demographics take the form of a bimodal distribution where the largest group of participants were in the range of $18 - 39$, with a mode

[2] i.e., time taken to press key 'J' on the keyboard for each user

in the 25 – 39 range, while many participants were also aged 60 or older. The distribution of race was heavily skewed toward the representation of whites, with most of the rest of the participants specifying Asian/Pacific Islander for race. Finally, a large majority of the participants specified that they were currently wearing corrective lenses, with smaller fractions indicating not wearing corrective lenses, either for having and not wearing or not needing them. Due to lack of space, we provide the plots of demographic distributions in Appendix A.

4.2 Hypotheses Tests

Results of the three Friedman's ANOVA tests corresponding to hypothesis H1, H2 and H3 are described below in Table 1. Histograms are included for each level of the variables of LDN, LDS, and ANG, grouped by the general parameter. The result of the hypothesis tests

Hypothesis	Parameter	χ^2	df	p-value
H1	ANG	0.594	4	0.9637
H2	LDN	110.73	5	2.2e-16
H3	LDS	57.43	3	2.08e-12

Table 1: Result of Friedman's ANOVA

for the Friedman's ANOVA for the parameters of both LDS and LDN had extremely low p-values, well below the standard significance measure of 0.05. Therefore, the null hypotheses that the distributions for each level of illusion spacing are identically distributed can be rejected, as can the null hypothesis that the distributions for levels of LDN are identically distributed. Observing Figures 3,4 corresponding to these tests further supports these rejections of the null hypotheses; the plot of LDS histograms shows differences in distribution height and spread throughout the histograms, and for the LDN histograms the differences are even more pronounced, with large differences in histogram height and the locations they are centered about.

In contrast, for variations in ANG, the resulting p-value is very high, indicating that the null hypothesis cannot be rejected. In this case there is little evidence to support the levels not being identically distributed. Examination of the histogram further supports this, as the distributions are observed to be clustered very closely together, with very similar height and spread, and centering in the same location.

For the multiple regression test corresponding to hypothesis H4, results were obtained for the significance of each independent variable level. The p-values obtained were low enough to comprise statistically significant results for the following categories: age group of 60+ (p-value 0.028358), race specified as white (p-value 0.017301), and the set of lens groupings (p-values of 1.64e-05, 1.31e-05, 9.01e-06, 2.69e-05, 1.01e-06, 2.02e-05, 7.70e-05, 4.68e-05, 0.005922, 3.22e-05, 0.000676, 0.006697).

The suitability of the multiple regression can also be further evaluated by plots obtained from the results of the regression. These include the Normal Q-Q plot of the data and the residual plot, which has residuals plotted versus fitted values. This Q-Q plot from the regression shows points with theoretical quantiles in the range of -2 to 1 clustered tightly around the diagonal, and although for values greater than 1 the points fall off the line somewhat, the general shape nevertheless demonstrates linearity of the quantiles against the residuals. The second plot, of the residuals against fitted values, shows a roughly spherical scatter of data points rather than a spread favoring a particular direction, and the spread above and below 0 y-value is nearly equal. This supports the assumptions of both linearity and equal variance. Thus, these two plots further bolster the claim that multiple regression was an appropriate test for analysis due to the satisfying of its assumptions. Due to lack of space, we provide the plots in Appendix C. Further, we perform

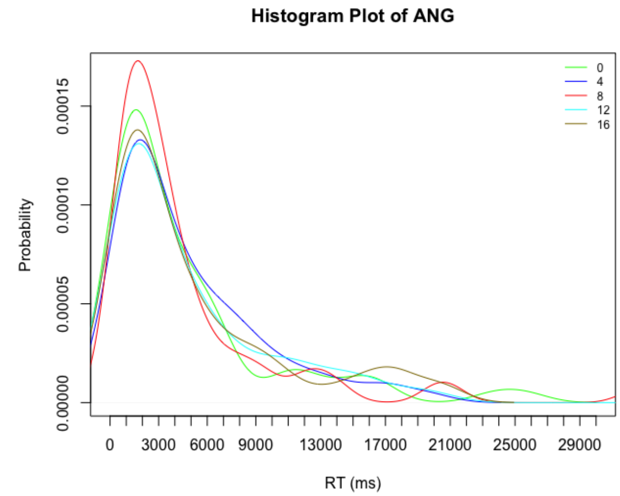

Figure 2: Histogram Plot of ANG

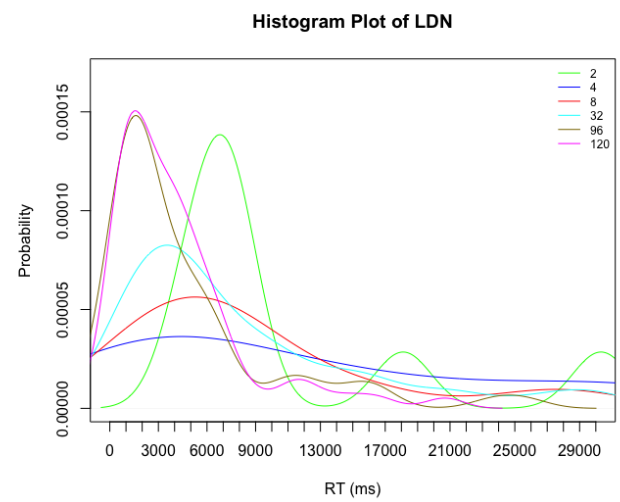

Figure 3: Histogram Plot of LDN

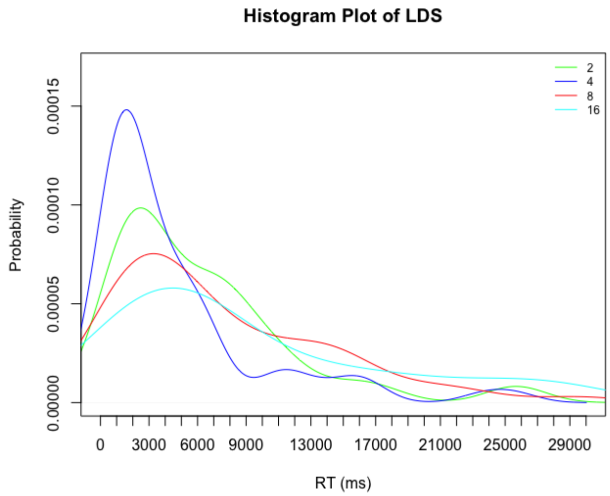

Figure 4: Histogram Plot of LDS

analysis on reaction time over various demographics parameters. Due to lack of space, we provide these plots in Appendix D.

5 DISCUSSION

Statistically significant results were obtained for the Friedman's ANOVA test with the dependent variable of space ratio of black and white lines, as well as with the dependent variable of line density. For both of these, the p-value alone was not tremendously expressive of the results that we obtained, and these can better be articulated by further discussion of the relevant histograms.

The combined plot with the density histograms displays a noticeable trend as the density value of the images increases. For densities of 96 and 120 lines, which were the higher numbers of lines before pixel aliasing, the histograms have the highest peaks and are shifted furthest left, indicating that participants could determine the illusion fastest for this category. The histograms for the other line numbers are shifted to the right, with this amount varying based on the count. The density of 32 lines is slightly lower, and the density of 8 and 4 are together lower still. Overall, the discernable trend is for lower densities being associated with longer times before the illusion is detected. There is one notable exception, in the histogram for density value of 2, which although centered furthest to the right does not fit the trend of being flatter, having a height almost as large as for 96 and 120. One way to account for this exception is by decomposing the histogram into contributions from different answers. This decomposes into two histograms, for "flickering motion" or "something else." The former gives a flat histogram more consistent with the combined histograms for the other densities, while "something else" contributes a large fraction of the histogram height. Therefore, understanding the reason for the "something else" answers could help address the discrepancy, although this would be left to future work, as we did not collect more detailed information.

The assembled plot of the space histograms also shows a trend among the levels of spacing. The histogram having the highest mode, the spacing of 4 (equal levels of black and white), is centered around the lowest time. Other proportions of black and white, such as the spacing 2 (doubled black) or 8 and 16 (doubled and quadrupled white, respecively), have increased time taken by participants and lower observation numbers, and so have progressively flatter histograms further to the right. Participants therefore appeared to be best able to detect the illusions with more equal color proportions.

Finally, regarding the multiple regression, although a number of variables yielded significant p-values in the regression, these do not all necessarily comprise meaningful results. One significant result that does not appear to be relevant is the signficance of specifying race as white. Due to this being an great majority of partcipants, that idenfication being associated with a low p-value does not necessarily generate insight. Conversely, the low p-value for ages above 60 does have a useful interpretation, as the boxplot for the reaction times of older participants is seen to have the largest spread by far, and a positive regression coefficient in the multiple regression results. Lastly, the categories of corrective lenses as a whole yielded highly significant p-values, and also have positive regression coefficients, leading to the interpretation that needing corrective lenses is generally positively associated with the dependent variable.

6 CONCLUSION

From the results, we observe that changes in density and line space ratio are related to how fast people observe motion in the illusions as well as how many people see movement at all. However, varying angle has no effect on reaction time. Our experiment encountered many limitations that we listed earlier. We focused on alleviating the effect of these limitations for future work. Although, our sample size was small, the results look promising for using this as a pilot to design and conduct a larger and stronger experiment with lessons learned during this project. We plan to continue this work under Dr. Fermüller and Dr. Mazurek's guidance.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Mazurek and Dr. Fermüller.

REFERENCES

[1] Michael Bach. 2004. Enigma: Visual Phenomena and Optical Illusions. (2004). http://www.michaelbach.de/ot/mot-enigma/

[2] Cornelia Fermüller and Henrik Malm. 2004. Uncertainty in visual processes predicts geometrical optical illusions. *Vision research* 44, 7 (2004), 727–749.

[3] Cornelia Fermüller, Robert Pless, and Yiannis Aloimonos. 1997. Families of stationary patterns producing illusory movement: insights into the visual system. *Proceedings of the Royal Society of London B: Biological Sciences* 264, 1383 (1997), 795–806. DOI: http://dx.doi.org/10.1098/rspb.1997.0112

[4] I. Leviant. 1996. Does 'Brain-Power' Make Enigma Spin? *Proceedings of the Royal Society of London B: Biological Sciences* 263, 1373 (1996), 997–1001. DOI: http://dx.doi.org/10.1098/rspb.1996.0147

[5] Isia Lviant. 1982. Illusory Motion within Still Pictures: The L-Effect. *Leonardo* 15, 3 (1982), 222–223. DOI: http://dx.doi.org/10.2307/1574685

[6] Xoana G. Troncoso, Stephen L. Macknik, Jorge Otero-Millan, and Susana Martinez-Conde. 2008. Microsaccades drive illusory motion in the Enigma illusion. *Proceedings of the National Academy of Sciences* 105, 41 (2008), 16033–16038. DOI: http://dx.doi.org/10.1073/pnas.0709389105

[7] Semir Zeki, John DG Watson, and Richard SJ Frackowiak. 1993. Going beyond the information given: the relation of illusory visual motion to brain activity. *Proceedings of the Royal Society of London B: Biological Sciences* 252, 1335 (1993), 215–222.

A APPENDIX: DEMOGRAPHICS

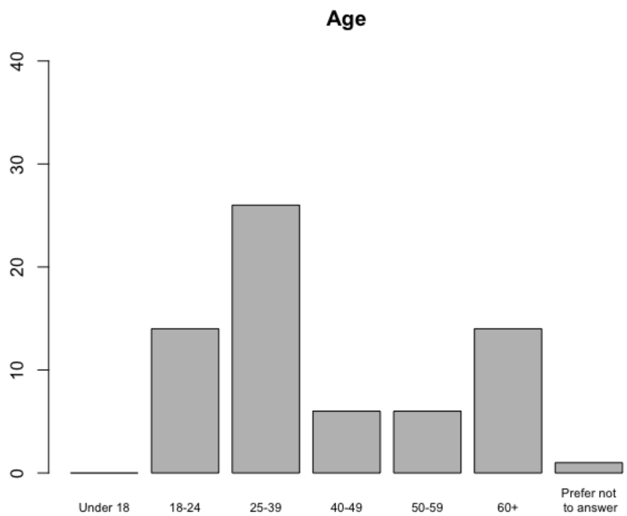

Figure 5: The age distribution is not as evenly distributed as we had hoped for. But it is also not heavily weighted only one one category.

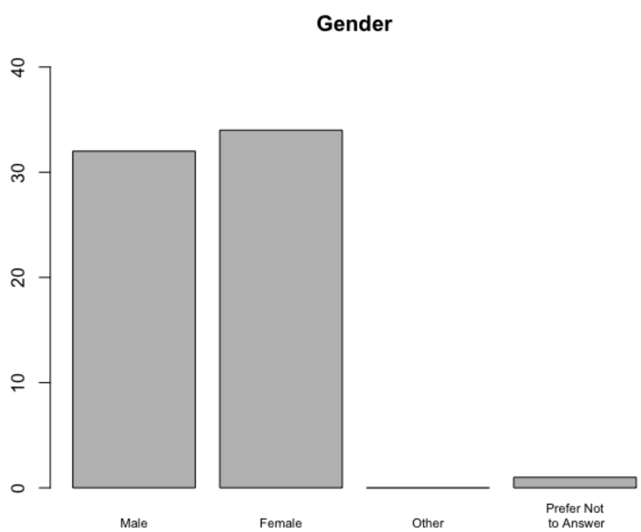

Figure 6: The Gender distribution has an approximate even split.

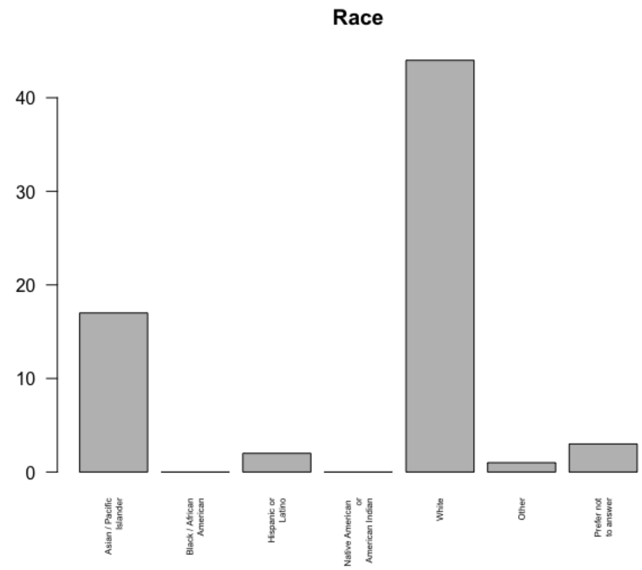

Figure 7: Race distribution is heavily weighted toward Whites, and about 1/3rd of that of Asians. Other races are not well represented.

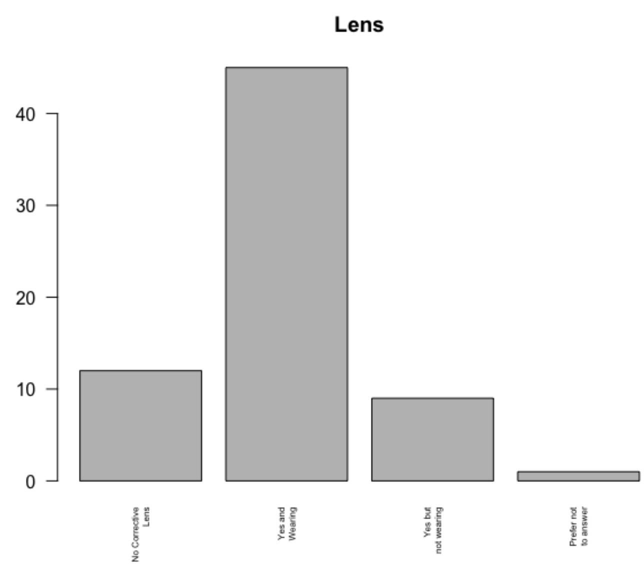

Figure 8: Optical defect and correction distribution is heavily distributed towards those who needs correction and were using the corrective lens.

B APPENDIX: ADS AND SOCIAL MEDIA

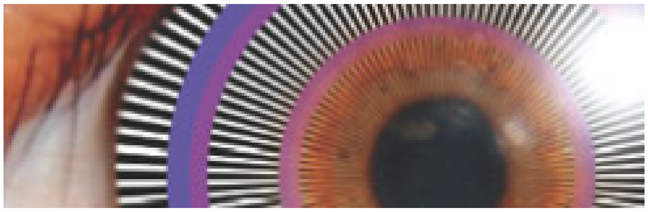

Figure 9: Survey Teaser Image

C APPENDIX: REGRESSION

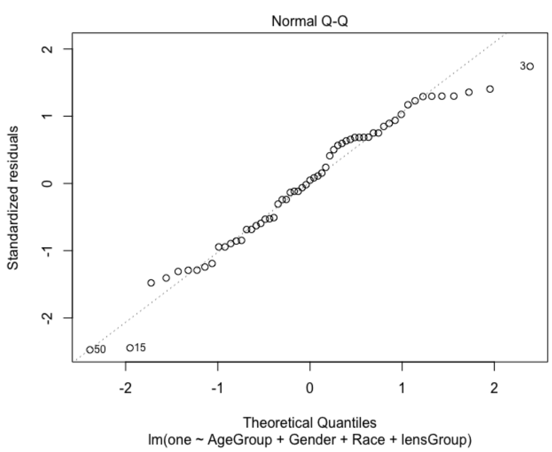

Figure 10: The Q-Q plot demonstrates approximately linear behavior.

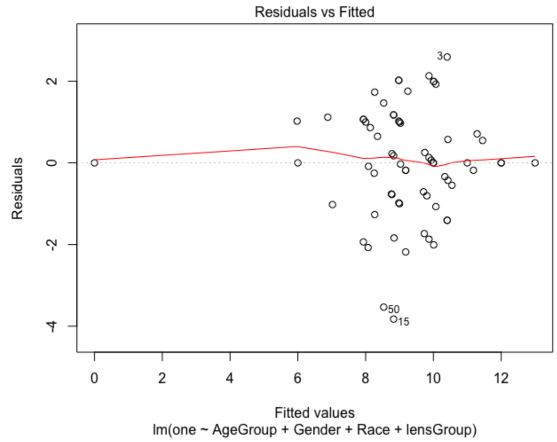

Figure 11: Symmetric residual distribution about the y-axis.

D APPENDIX: ADDITIONAL ANALYSIS

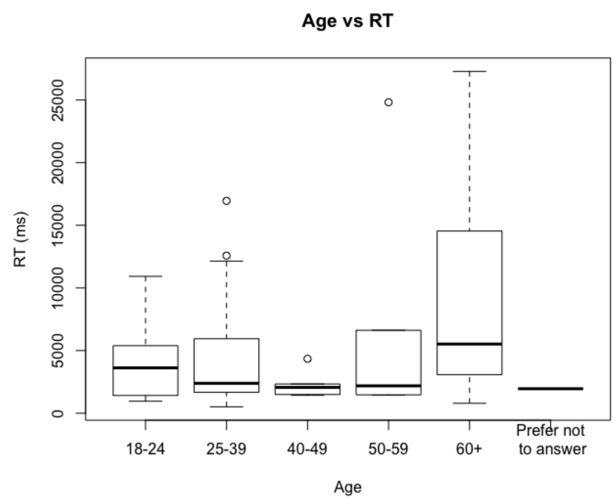

Figure 12: There seems to be a sudden inconsistency for the 60+ age group.

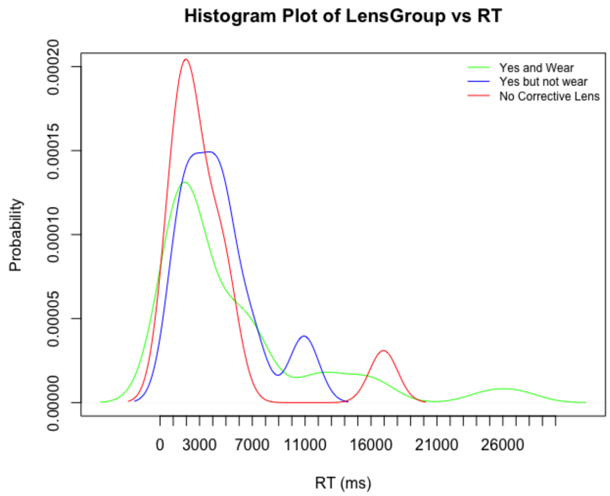

Figure 13: No difference between people without defects & ones wearing lens. Takes longer for people with eye defects not wearing lens.

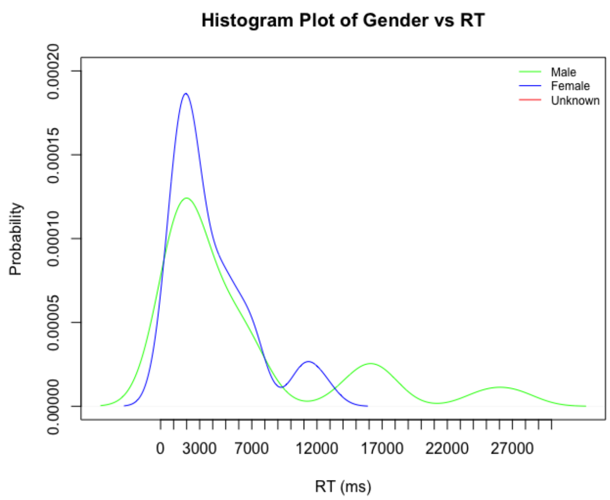

Figure 14: No difference between time taken to see the illusion across gender.