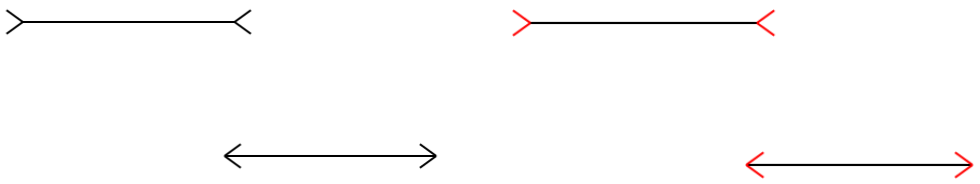
Matt Burns & Ryan Christensen
CS5890
12/07/2024

Final Paper - Muller-Lyer Replication and Figure Color

Introduction

The Muller-Lyer is a well known optical illusion consisting of a set of arrows, when views of the illusion are asked to determine which of the two arrowed lines in the illusion are of greater length viewers tend to indicate that one of the lines is longer than the other. However, in a standard Muller-Lyer illusion illustration, the two lines will be of the same length. These illusions were initially devised by the German sociologist Franz Carl Muller-Lyer, the effect has been well recognized and studied since the initial observation by Muller-Lyer.

Muller-Lyer illusions can appear in a number of configurations and variations, with the effect being observable in each to a greater or lesser extent. This study focuses on the classic configurations of the illusion. A representative illustration of the two versions of the Muller-Lyer illusion as presented in this study can be seen below (the lengths of lines in both of these figures are identical). More information on how these figures were varied in our experimental setup can be found in our methods section.



Research on the Muller-Lyer illusion has focused on determining the source of the illusion, as well as on whether the illusion can be unlearned through regular structured practice. This aims to determine through a replication study whether the illusion can be effectively unlearned through both intensive and intermittent practice. We also aim to determine whether arrow color has a measurable effect on the accuracy of a viewers estimation of the length of Muller-Lyer figures.

Report Repository: *https://github.com/m6urns/cs6890-final-project-muller-lyer*
Project Repository: *https://github.com/m6urns/muller-lyer-replication*
Project Slides:
https://docs.google.com/presentation/d/1DmQgoHJNB9nke2X_qtg61wD-CzDcroaIA-8bzbfaBbM/edit#slide=id.g31be368f937_0_14

Literature

The initial impetus for this replication study and study on the effect of color was a review of the paper titled “The Effect of Practice on the Perception of the Muller-Lyer Illusion”, written by E.O. Lewis and published in 1909 [1]. This paper represented an early effort to understand the cause of the illusion and in addition to exploring whether the effect could be unlearned through practice. The paper also made an effort to determine whether the effect could be caused by the physical configuration of the eyes. This paper experimentally determined that the effect could be unlearned through practice, and the author believed that the effect was not as a result of physical characteristics of the eyes of the viewer, and was rather psychological in nature.

Further research has been conducted on the illusion attempting to determine its source. Some studies have pointed to environmental factors as being the source of the illusion and reported a lack of the effect in individuals living in less developed environments [2]. This is referred to as the perspective explanation for the Muller-Lyer illusion, and is one potential explanation.

Another potential explanation to the occurrence of the illusion is the centroid hypothesis, this explains the illusion as a result of the neural process of centroid extraction. Centroid extraction is related to a process called spatial pooling of positional signals. This pooling of signals from neighboring objects has the advantage of fast, reliable visual processing at the loss of some positional acuity. This loss of acuity as a side effect of more efficient visual processing is hypothesized by proponents of the centroid theory to account for the Muller-Lyer effect [3].

Follow-ups to the E.O. Lewis unlearning studies have explored both the effect of previous training and have re-confirmed the assertion that practice can improve an individual's ability to see through the effect [4].

Methods

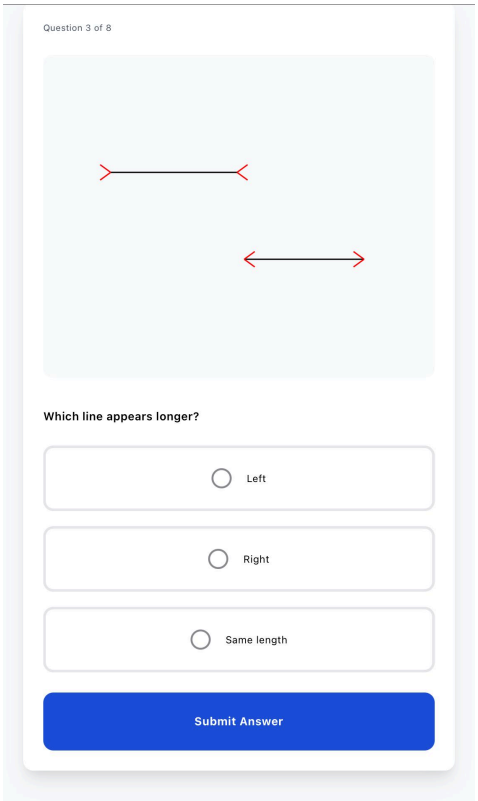
Data Collection

Data was collected through a custom web frontend developed in React JS with a Python Flask backend. Figures were generated programmatically as SVG files. We used the following settings to generate figures for our intensive practice group: “generation/illusion-generator.py --days 20 --num_illusions 25 --arrow_color "red" --generate_duplicates --generate_control”. Note that for our intensive group these settings produced a daily set of 100 images for participants to train on, an initial set of 25, duplicated into a black and a red version. These two versions were administered in identical but mixed “Fast” and “Slow“ variants with the “Fast” version allowing the user to review the images for 50 ms and “Slow” version allowing the users to review the images for up to 500ms.

Similar to our intensive group we used the following to generate illusions for our intermittent group: “generation/illusion-generator.py --days 20 --num_illusions 8 --arrow_color "red" --generate_duplicates --generate_control”. These settings yielded a total of 32 images for users

in the intermittent group to review and practice on each session, in “Fast” and “Slow” variants of 16 images each. Settings for this group beyond the reduced number of images were identical. Twenty images were produced for this group for simplicity, however only the first seven sessions of generated images were used. The script used to generate SVG images with these settings can be found in our project repository.

At the time of generation the pixel length difference between the two figures was varied, and this information was recorded in a metadata file for each day, this metadata is also available in our repository. In addition to an experimental group where the length difference between figures was varied an additional “Control” group was produced, in this group the only difference from the experimental group was that there was no variation in the lengths of figures, in this group all figures presented were of identical length, although the illusion was still present due to the nature of Muller-Lyer figures.



Once image sets had been produced we utilized the generated images and metadata to produce a set of daily illusion practice “quizzes”. These “quizzes” were served to users though our custom frontend. This allowed for simple administration, allowing users to participate by simply scanning a QR code or navigating to a link. At this point users were asked to input their selected user name and select the relevant daily “quiz” from a drop down list. As users worked though each quiz they were presented with three options with the image of the illusion in question displayed above the answers. If the user took longer than the allotted time per image to evaluate, the figure disappeared, however the users could still select a response. The time the user took to provide a response was also recorded. A screenshot of the interface that users interacted with during the “quiz” can be seen to the right. Users were presented with a single question for each prompt for all versions of the “quiz”: “Which line appears longer?”. The user was then provided with three possible responses: “Left”, “Right”, or “Same length”. Users were required to select a response before moving onto the next question.

Data was collected in two separate groups: an intensive group (n=4) and an intermittent group (n=13). The intensive group was recruited opportunistically and the intermittent group consisted of individuals from the Fall 2024 Data Science and the Brain class. Individuals in the intensive group were asked to take two of these “quizzes” each day for 20 days, they were not instructed on when or where to take these quizzes so settings may have varied. The intermittent group was collected at the beginning of each class session over the course of approximately 20 days; this resulted in a total of seven sessions collected in the intermittent group. Data was collected separately for each group, data was then aggregated into a single CSV file with additional contextual information relating to figures, whether a user's answer was correct, this data was derived from the metadata

generated at the time the SVG figures were produced. Exact information on this processing step can be found in the processing section of our project repository.

Data Cleaning

Once we had concluded data collection from both our intermittent and intensive groups we cleaned and validated our collected data. This cleaning and validation step was done through a series of Python scripts embedded in a Jupyter notebook. Shortcomings of our user interface made it possible for users to take quizzes more than once, as well as use slightly incorrect user names. We took steps to identify instances where users had taken a quiz multiple times, had not finished a quiz or had used the incorrect user name. In cases of multiple quizzes or incomplete quizzes we removed those responses. We corrected incorrect usernames as needed. Removals due to duplications and incomplete quizzes were minimal. We also dropped some users from our intensive group who had not completed the entire 20 day course. Users in the intermittent group who had not been present to complete more than 50% of the administered sessions were also dropped. In both groups this represented a removal of two individuals from each group.

Data Analysis

We performed our data analysis for this study in Jupyter notebooks, utilizing a standard set of Python based data science analysis tools. Specifics on which tools and libraries and how they were used can be found in our project notebook. Results were analyzed for significance through standard T tests, with a standard p value of 0.05.

Results

Effect of Intermittent and Intensive Practice

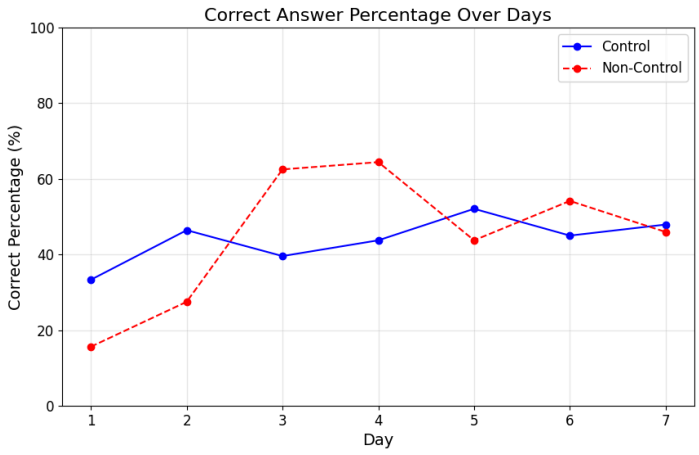
After having run the studies we analyzed the results in various ways. One of the main analyses we ran was between the difference in performance (ability to bypass the effect of the illusion and answer correctly) in the intermittent group and the intensive group.

The intermittent group showed significant improvement over the course of the study with a 30.4% improvement, from a 15.6% average on day one to a 46% average on day seven with an average of 44% across all days, with their biggest increase in performance being in the first three sessions. The control group in the intermittent group didn't show significant gains in performance with an average of 44% throughout the course of the study.

The intensive group showed a slight decrease in performance over the course of the study with a 16.5% decrease, from a 54% average on day one to a 37.5% average day twenty, with a 43.7% average across all days. This decrease could be explained by a number of factors, most up front of all is the small sample size. Given that the sample size was only four for the intensive group there is a lot of variability and sway given the performance. However, there was a 24%

increase in performance of the control group, from a 46% performance on day one to a 70% performance on day twenty with an average of 62.7% across all days.

Effect of Pauses Between Sessions



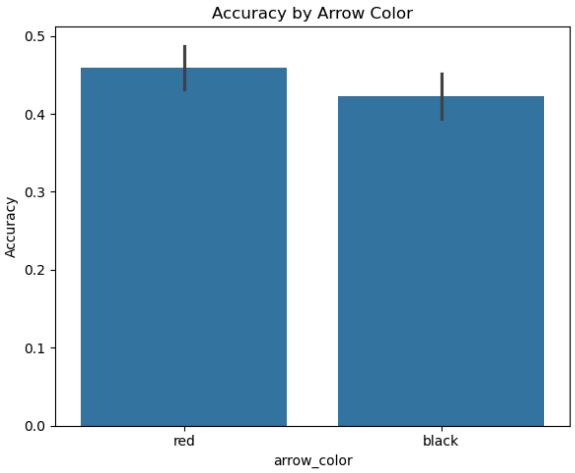
An interesting finding that stood out while interpreting the data was the potential effect of pauses between sessions. Focusing on intermittent group, as they were only group with given breaks, we found that every successive trial within the same week saw an increase in performance, while each pause between sessions due to a weekend or otherwise saw a decrease (apart from day one to day two).

Day two to day 4 is a successive Monday, Wednesday, then Friday of intermittent class period trials. This is where we see the primary growth of the group with a 36.9% increase in performance. Then there is a weekend break in between days four and five, in which the performance dropped from 64.4% to 43.75%. Following the significant drop there is another successive intermittent class trial, with an increase in performance of 10.35%. Then there is a significant break of five days in which the average performance drops 8.2%.

The observed improvement in performance during successive intermittent trials, along with the noticeable decline in performance following a pause between sessions, provides some evidence that supports the idea of unlearning in relation to the seeing through the illusion. This pattern suggests that time spent away from exposure to the illusion may have a measurable impact on performance outcomes. Specifically, the data indicate that breaks or pauses in engagement with the illusion could contribute to a temporary reduction in the ability to perceive the true lengths of the lines, and that consistent successive viewing of the illusion could improve the ability to see through the illusion.

Effect of Arrow Color on Accuracy

One of the main goals in this study was to compare the performance accuracy between different arrow colors to see if there is a difference in the illusions effect based on the colors. Given the average accuracy across the study we see that the results show a slight favor of the colored (red) arrows in comparison to the non-colored (black) arrows, also the results were not statistically significant with a p-value of 0.076.

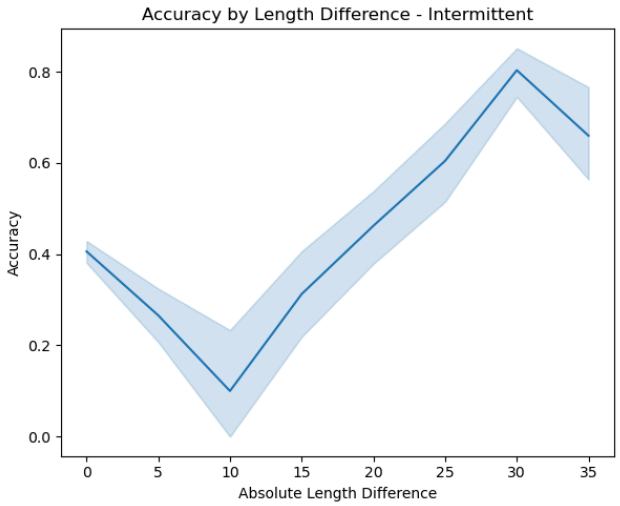


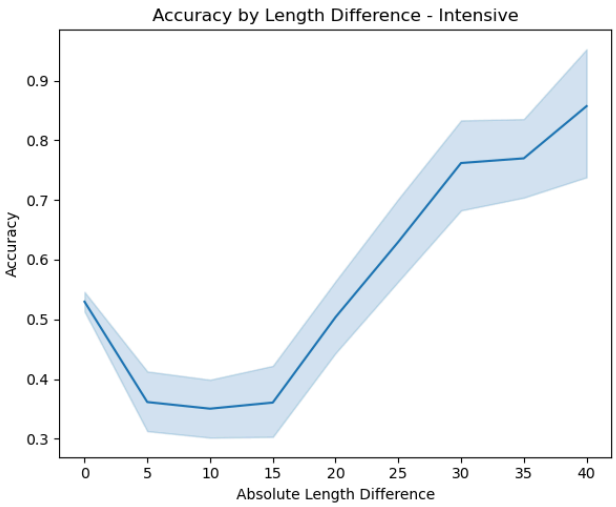
The change in color has a possible effect on the learning/unlearning in relation to seeing the illusion. This would have to be studied further to show any conclusive evidence.

Accuracy Increases with Difference in Figure Length

We observed in both of our groups that there was a significant correlation in participants' ability to differentiate figure differences as the difference in figure length increased (Correlation: 0.211, p-value: 0.000). Within our intermittent group of participants we observed that individuals were least able to differentiate when the absolute difference between the two lines in a figure had an absolute difference of 10 pixels. This accuracy increased as the absolute difference between the figures increased from 10 pixels. Interestingly it seemed that participants were more able to differentiate between figures with an absolute difference of 5 pixels than those with an absolute difference value of 10.

We observed a similar although weaker trend in our intensive group (Correlation: 0.068, p-value: 0.000), we do note that this group was smaller than the intermittent group. We observed that, similar to the intermittent group, individuals in the intensive group



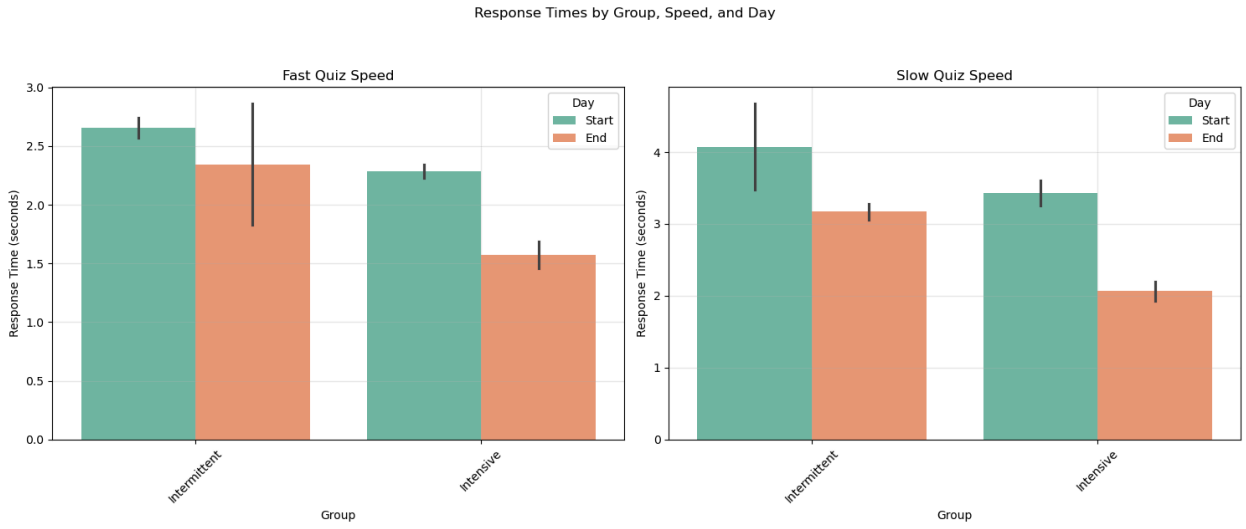


were least able to differentiate at around a difference of 5-15 pixels, after which the participants' ability to differentiate increased.

Response Time Decreases from Initial to Final Sessions

We observed that the response times for both fast speeds and slow speed (the time a participant took to select an answer in the web interface) decreased from the beginning to end of the experiment. This was observed in both intermittent and intensive groups and as mentioned at both “Fast” and “Slow” versions of the “quizzes”. Participants still took more time on average to examine the figures in the “Slow” version of the “quizzes”, but the amount of time that individuals took to respond to the question of which figure was longer was very similar between “Fast” and “Slow” speeds.

We are unsure if participants reduced the amount of time they spend examining figures due to study fatigue or increased confidence in their estimates. It is notable that in our intermittent study group that accuracy increased significantly from the first day of practice to the final day of practice. Contrary to expectation the intensive experimental group decreased in performance from beginning to the end of the study. Participants in the intensive group communicated that by the end of their twenty days of practice they were extremely fatigued, in addition several individuals dropped out of the intensive study group before completing the full duration. This would seem to suggest a potential mixed cause for this decrease. More inquiry into the root cause of this decrease would be needed for further clarity.



Discussion

This study replicated and expanded upon previous research on the Muller-Lyer illusion, focusing on two primary objectives: whether the illusion can be mitigated through structured practice and whether arrow color affects perception accuracy. The findings suggest that while intermittent practice significantly improved participants ability to overcome the illusion, intensive practice showed inconsistent results, likely due to the small sample size. Also, arrow color did not yield statistically significant differences in performance accuracy, though there was a slight trend favoring colored (red) arrows.

The results also highlighted the impact of session spacing on performance. Breaks between sessions led to declines in accuracy, whereas successive sessions saw notable improvements, supporting the idea that inconsistent exposure creates an unlearning effect. The study also showed that performance accuracy improves as the length difference between lines increases.

Citations

1. Lewis, E. O. (1908). The effect of practice on the perception of the Müller-Lyer illusion. *British journal of psychology*, *2*(3), 294.
2. Letourneau, J. E. (1976). Effects of training in design on magnitude of the Müller-Lyer illusion. *Perceptual and Motor Skills*, *42*(1), 119-124.
3. Whitaker, D., McGraw, P. V., Pacey, I., & Barrett, B. T. (1996). Centroid analysis predicts visual localization of first-and second-order stimuli. *Vision Research*, *36*(18), 2957-2970.
4. Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception* (Vol. 310). Indianapolis: Bobbs-Merrill.

Appendix

Table 1 - Summary of Accuracy and Response Time for Intermittent Practice

Group	Quiz Speed	Day	Accuracy (Mean)	Response Time (Mean)	N Trials
Experimental	Fast	Start	0.194	2.57	160
Experimental	Fast	End	0.453	1.83	128
Experimental	Slow	Start	0.109	3.48	128
Experimental	Slow	End	0.484	3.06	128
Control	Fast	Start	0.313	2.74	48
Control	Fast	End	0.469	2.86	32
Control	Slow	Start	0.354	4.67	48
Control	Slow	End	0.406	3.28	32

Table 2 - Summary of Accuracy and Response Time for Intensive Practice

Group	Quiz Speed	Day	Accuracy (Mean)	Response Time (Mean)	N Trials
Experimental	Fast	Start	0.521	2.34	144
Experimental	Fast	End	0.375	1.46	48
Experimental	Slow	Start	0.563	3.60	144
Experimental	Slow	End	0.375	1.93	48
Control	Fast	Start	0.547	2.23	48
Control	Fast	End	0.620	1.69	50
Control	Slow	Start	0.460	3.26	50
Control	Slow	End	0.780	2.20	50