DETERMINING THE THRESHOLD BETWEEN TWO NUMBERS

by

DANIEL JAMES HRUSOVSKY

A manuscript submitted to the Faculty of John Carroll University

in partial fulfillment of the requirements for

graduation with a Bachelor of Science of Mathematics

2025

Hrusovsky, Daniel James (Bachelor of Science, Mathematics)

Determining the Threshold Between Two Numbers

ABSTRACT

Writing is a unique style of language that uses the perspective of the reader and writer to determine connections in figures. Letters can present themselves different to each individual, creating a threshold. The aim of this research is to determine the threshold between two numbers; 7 and 1. This will be done through a scaling system created by matrix multiplication and survey sequencing using a k-randomness algorithm. This research will look at two different formats of the numbers; first without an additional horizontal line and then with a line. Analysis will be done on both formats by looking at the threshold distribution to determine the threshold. The results for the first format concluded that most individuals determine the threshold to be skewed right, which denotes more individuals view the figures as 1 rather than 7. For the second format, the distribution demonstrated large frequencies of 7, where more individuals view the figures as a 7 rather than 1.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Handwriting has been a formidable transferring of information for over 5,000 years. Since the systematic form of writing has existed, it has created a unique way to provide facts and details about any concept. However, handwriting's flaw is in the interpretation of the reader. In Jaroslaw Moszczynski's *The multi-individuality of handwriting*, he writes, "The appearance of handwriting may be subject to temporary or permanent changes resulting from various external and internal factors" [5]. While individuals conjure letters together to create words, some get lost in translation. This is because of a natural disconnection between the reader and writer.

Writing is a unique, personalized experience that can change on a daily basis for each person. This is commonly seen when attempting to read a note you have previously written quickly. Despite being your own handwriting, it is still difficult to read. Whether it is your own handwriting, a doctors note, or a colleagues homework, these differences can cause struggles in interpretation.

Does this mean, however, that it is the reader or the writers fault for this miscommunication? One could argue the writer since their writing was too poor. The other could debate that the reader is at fault due to their lack of understanding what has already been constructed. In actuality, neither are at fault. This is because each individual has their own understanding for each character. Moszczynski states, "The manner of writing formed in such a way, remains relatively stable, till late life, when it becomes subject to characteristic deformations" [5]. While there is a "perfect" way to write characters, the imperfection of humans restricts us from consistently writing this way. Thus, alternatives styles arise.

Thus, for each person, there is a "threshold" for each symbol, where they believe one figure reigns supreme over another. There is an instance that switches the capitalization for the reader. For example, Figure 1.1 shows an uppercase T followed by a lowercase t.

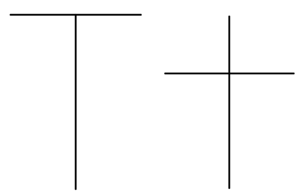

Figure 1.1: An uppercase T on the left with a lowercase T on the right.

While it is no guarantee, most individuals would believe the figure on the left to be an uppercase T while the right figure to be a lowercase t due to the positioning of each line. Since the left figure's horizontal line is at the very top, it would fit into most's idea of uppercase rather than lowercase. This is similar with the right figure's line, as it is closer to the middle than it is the top. However, if we were to move this horizontal line to a spot between the top and middle, it is less apparent whether it is uppercase or lowercase.

There is assumption that some individuals believe Figure 1.2 is uppercase, while others believe it is lowercase. Thus, there is a threshold of belief for whether the letter has capitalization.

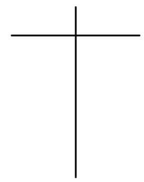

Figure 1.2: A figure with a horizontal line intersecting meant to represent either an uppercase or lowercase T.

We will be looking at a similar example to determine the threshold between two numbers; 7 and 1. To do this, we will view former research on the perspective on color. This research, created by Patrick Menault, looked at the threshold between colors green and blue. This was done by responses through the website *ismy.blue*. Individuals would be given a color between green and blue and respond depending on which color most aptly fit [4]. This would be done until the color's shown would converge to a point where everything to the left is green and to the right is blue. This is shown in Figure 1.3, where each square represents the sequential colors pictured to the respondent. This research will help us determine our procedure to discover the numerical threshold.

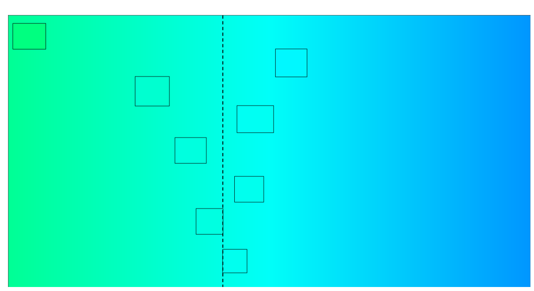

Figure 1.3: A simulation of *ismy.blue* with colored squares to represent the color of each further question.

CHAPTER 2

METHODOLOGY

A scale is necessary to determine the threshold between the two numbers. Without a scale, it would not be possible to denote differences between figures. For the research regarding color, an HSL scale is used. This scale uses hue, saturation, and lightness to assign colors numerical values. It uses a scale of 150 to 210, where 120 is true green and 240 is true blue. Figure 2.1 depicts the HSL scale used.

Figure 2.1: The HSL scale from 150 to 210.

For a change in appearance of numbers, there is no designated scale. Thus, one will need to be made by finding the appearance of our true values. In our case, we want a scale that moves from a true 7 to a true 1.

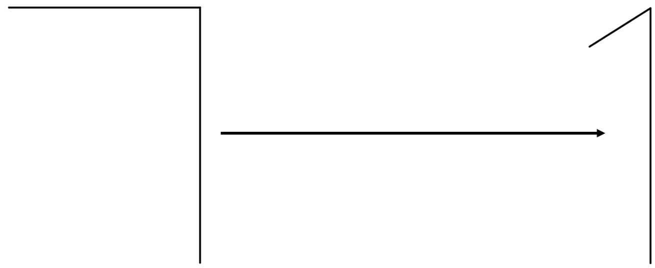

Figure 2.2: A 7, left, with an arrow pointing to a 1, right.

Notice how the difference between these two numbers is through the "extended" line, which is long and horizontal for the 7, is short and diagonal for the 1. Thus, we need to create a scaling mechanism to mathematically initiate this alteration. This is done through linear algebra.

2.1 Linear Algebra

Linear algebra is a branch in mathematics that deals with using vector fields and matrices to discover solutions of linear equations. With this, it uses systems of linear equations to determine relationships of variables through matrices. These matrices can alter these vectors through a multiplication system. Using linear algebra can serve multiple real-world purposes, primarily through visual changes in graphics or through statistical prediction models.

Matrix transformations are geometric transformations of vectors through multiplication that creates an altered, or transformed, vector space [1]. It creates the ability to modify the displacement, direction, rotation, and dimensional space of the vector. It does this by altering the coordinate space that the vector resides in rather than the vector itself.

We will use matrix transformations to move the "extended" line of the figure. The extended line is the only part of the figure that needs to move in order to make the transfer possible. To move this extended part of the number, we can use a *shearing transformation*. Sheared transformations, or shear mappings, allow parallel displacement in an axis while maintaining the position of all other axes [3]. This shearing keeps the proportion of all distances from a point to an axis. Figure 2.3 shows how the horizontal shearing transformation changes upon the coordinate plane in the positive direction.

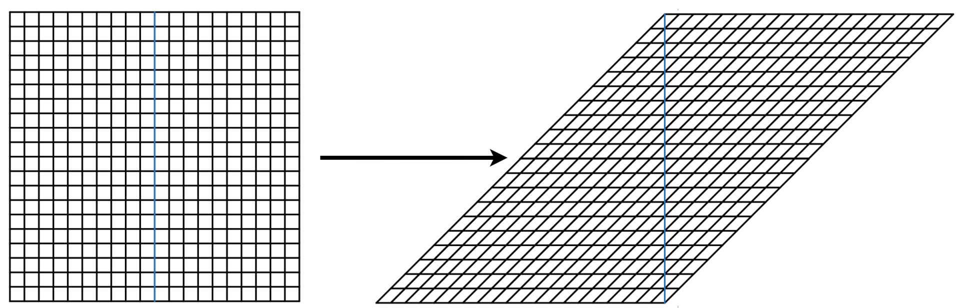

Figure 2.3: A coordinate plane sheared in the positive horizontal direction in black with the y-axis highlighted in blue.

As the transformation occurs, the plane does not move. Rather, the plane is moving about the axis. With this, as the plane moves, the coordinates shift laterally, as it is a horizontal shearing. This transformation can be done vertically, as well. As the vertical shear occurs, it performs similarly to the horizontal transformation, except it moves longitudinally rather than laterally, as shown in Figure 2.4.

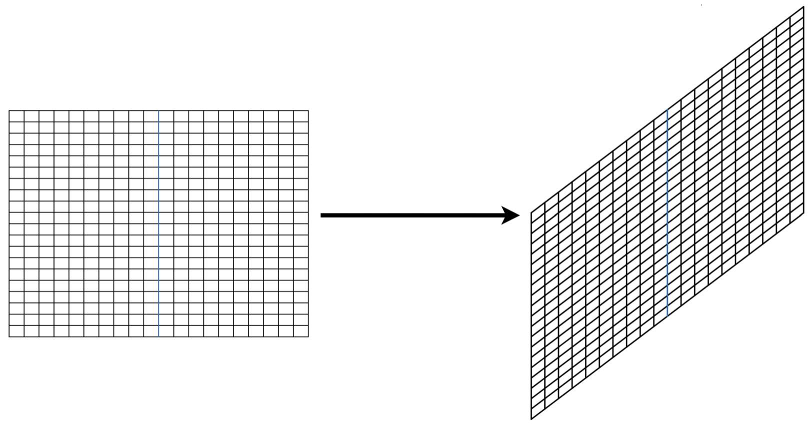

Figure 2.4: A coordinate plane sheared in the positive vertical direction in black with the y-axis highlighted in blue.

By using a vertical shear, we can move this extended line diagonally. This will allow the transition from the straight, horizontal line of the 7 to the diagonal line of the 1. Mathematically, we can represent this shearing transformation using matrices. The equation for a vertical shearing transformation on a two-dimensional vertex is defined as:

$$\begin{bmatrix} 1 & 0 \\ g & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_1 \\ gx_1+y_1 \end{bmatrix},$$

where x_1, y_1 is a point on the xy-plane and g is the magnitude of the shearing. For example, if we were to perform a vertical shear transformation on the point $(2,3)$ with a shearing magnitude of $\frac{3}{2}$, the transformed vertex would be,

$$\begin{bmatrix} 1 & 0 \\ \frac{3}{2} & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ \frac{3}{2}(2)+3 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}.$$

When looking at this graphically, it is seen how this change works on a singular point.

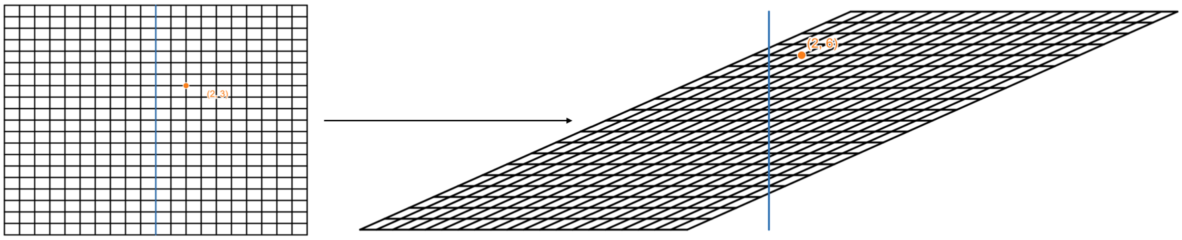

Figure 2.5: A coordinate plane with the point $(2, 3)$ being transformed to $(2, 6)$ after the shearing transformation.

While the coordinate remains in at the point $(2, 3)$ in the transformed vector space in Figure 2.5, it has moved in the standard plane to $(2, 6)$. This is a concern when creating the equation, as the point needs to stay constant while the shearing is occurring. This can be solved by using a vertical displacement matrix. The equation for a vertical displacement matrix on the point (x_1, y_1), where (x, y) is the desired point:

$$\begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix},$$

where c is the factor of displacement. By solving for c,

$$c \cdot (y_1) = y$$
$$c = \frac{y}{y_1}.$$

By substituting the sheared vertex,

$$c = \frac{y}{y_1}$$
$$c = \frac{y}{gx_1 + y_1}.$$

The desired point is found through the product of the displacement and shearing matrix on a vector. Thus,

$$\begin{bmatrix} 1 & 0 \\ 0 & \frac{y}{gx_1+y_1} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ g & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{y}{gx_1+y_1} \end{bmatrix}\begin{bmatrix} x \\ gx+y \end{bmatrix} = \begin{bmatrix} x \\ \frac{y}{gx_1+y_1}(gx+y) \end{bmatrix}.$$

Next, an equation is needed to pass through this point. This can be done using a point-slope equation. The equation of a line in point-slope form is:

$$(y - y_1) = m(x - x_1),$$

where (x_1, y_1) is a point and m is the slope of the function. Thus, we need a point and a slope. The point will be the aforementioned point $(x, \frac{y}{gx_1+y_1}(gx+y))$ where $x = x_1$ and $y = y_1$, our initial point. The slope that will be used is $m = \frac{gy_1}{gx_1+y_1}$, as this creates a perfect shaping for the figure. Thus, through substitution,

$$\begin{aligned} y - y_1 &= m(x - x_1) \\ y - (\tfrac{y}{gx_1+y_1}[gx + y]) &= \tfrac{gy_1}{gx_1+y_1}(x - x_1) \\ y - (\tfrac{y_1}{gx_1+y_1}[gx_1 + y_1]) &= \tfrac{gy_1}{gx_1+y_1}(x - x_1) \\ y - y_1 &= \tfrac{gy_1}{gx_1+y_1}(x - x_1) \\ \tfrac{gx_1+y_1}{y_1}(y - y_1) &= g(x - x_1) \\ (\tfrac{gx_1}{y_1} + 1)(y - y_1) &= g(x - x_1). \end{aligned}$$

This is the equation to create the desired figure. Figure 2.6 shows the figure with this newly acquired figure as g's magnitude increases.

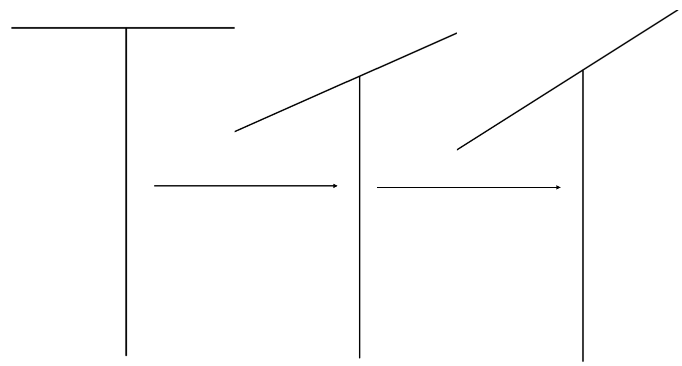

Figure 2.6: The figure of the equation at $g = 0$, left, $g = 1$, middle, and $g = 3$, right.

The next step is to bound the extended figure's length. Let x_2 be the endpoint of the extended figure and t be an arbitrary positive integer. Thus, we can create a boundary for this extended line:

$$x_2 + g^t \leq x \leq x_1.$$

Thus, our equation is $(\frac{gx_1}{y_1}+1)(y-y_1) = g(x-x_1)$, where $x_2 + g^t \leq x \leq x_1$. Below is the same intervals of the equation with these new boundaries.

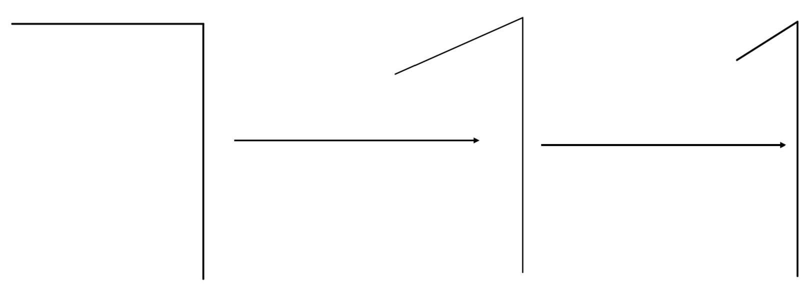

Figure 2.7: The figure containing boundaries at $g = 0$, left, $g = 1$, middle, and $g = 3$, right.

This will be the first figure used to determine a threshold between the two numbers. However, we can use a similar figure to see how slight changes in numbers can play a role in determining thresholds.

2.2 Second Figure

Not everyone writes their 7's in the same manner. Many people include a horizontal line in the middle. This is to help differentiate a 7 to similar figures.

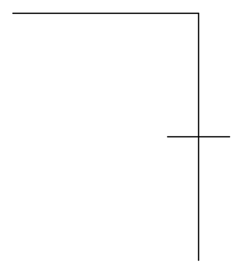

Figure 2.8: The number 7 with a horizontal line in the middle.

In Figure 2.8, the horizontal line is portrayed in the middle of the line. Similarly, some individuals write their 1's with a horizontal line, but rather towards the bottom, as shown in Figure 2.9.

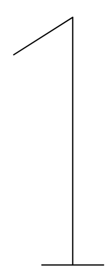

Figure 2.9: The number 1 shown with a horizontal line at the bottom.

Thus, we can find a similar threshold between these two figures using the same equation and boundary with an additional horizontal line intersecting the vertical line. Let x_3 be the initial point of intersection between the new horizontal line and the vertical line, while ϵ is the distance between x_3 and the vertical line's endpoint. Thus,

$$y = x_3 - \tfrac{g}{\epsilon} x_3,$$

where $x_1 - \frac{1}{2} \leq x \leq x_1 + \frac{1}{2}$. This results in our second figure, which is shown below in Figure 2.10.

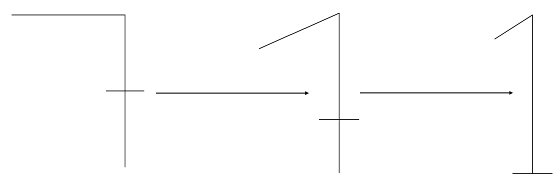

Figure 2.10: The second figure at $g = 0$, left, $g = 1$, middle, and $g = 3$, right.

These are the two figures that will help create a scale to determine the thresholds between 7 and 1.

2.3 Survey Research

In order to understand how humans respond to these figures, a presentation method is necessary. Thus, a survey is an applicable and feasible method to receive eligible, unbiased data. However, the survey cannot contain a full scale from 7 to 1, as the survey would take too long to answer and answers would be too minuscule to determine accurate results.

As a result, the survey will have 13 images for each figure. The response from the individual will determine if they believed the figure is a 7 or a 1. No additional information was asked about the figure or the individual. Each image is the aforementioned figures at different magnitudes of g. These values range from 0 to 3 at quarter intervals. Thus, the values of g are as follows:

$$G = \{0, \tfrac{1}{4}, \tfrac{1}{2}, \tfrac{3}{4}, 1, \tfrac{5}{4}, \tfrac{3}{2}, \tfrac{7}{4}, 2, \tfrac{9}{4}, \tfrac{5}{2}, \tfrac{11}{4}, 3\},$$

where G is the set of values for g. Figure 2.11 shows how these values of g creates a scale similar to the HSL scale used in the threshold research of color.

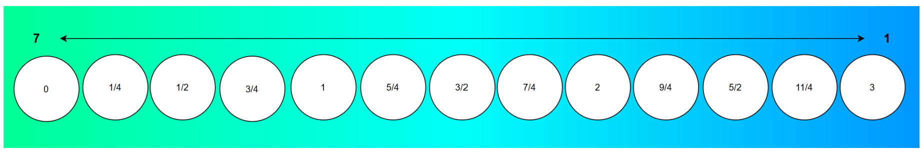

Figure 2.11: Values of g increasing from left to right in front of a color spectrum of green to blue.

As the value of g increases, the intention is for most responses to deviate from choosing 7 and to 1, similar to the intention that most individual will see blue as the HSL scale increases in value. This can be seen in Figure 2.12 with the thirteen displayed images for both section one and two depicted on the former HSL scale.

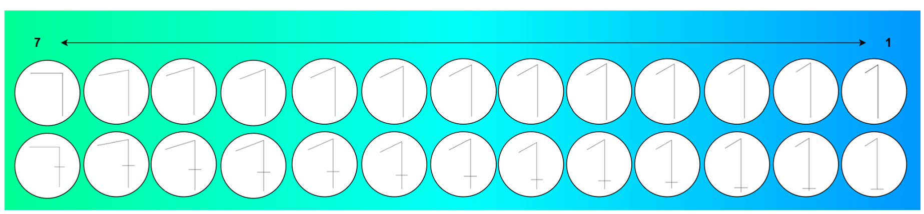

Figure 2.12: Numerical figures based on values of g from left to right in front of a color spectrum of green to blue.

This survey consists of two different sections. The first portion contains images of the first figure, while the second section contains images of the second figure. These questions are partitioned into four different blocks, with the first three containing 3 questions and the final block containing 4 questions. However, these blocks cannot show a linear progression of g, as it can show biased behavior from the respondent. Revealing these images in order would allow the individual to manually determine their threshold rather than determining it through from the figures.

To prevent biasing, the questions will be feel randomized to the testee. Question randomization prevents order from being a factor and minimizes survey fatigue [2]. Thus, creating a process to "randomize" the survey will help utilize the true difference in perceptions.

To create an unbiased, randomized feeling, the survey will be using an altered *k-randomness* approach. The k-randomness estimation, or Martin-Löf randomness estimation, is an algorithmic random sequence that takes a set of k centroids, or the midpoint of a geometric subject, to create clustering for a group of data [6]. It takes organic data and classifies them based on the distance to each centroid.

To use the k-randomness algorithm, the first rule is to assign initial centroids of the set. This is to help guide the randomness of the clustering. To do such, a set of n observations is provided. The algorithm attempts to partition these n observations into designated sets. After this, it attempts to classify the data into clusters. Once this is done, it finds the new centroids based on the previous iteration and repeats this process until there is convergence. Mathematically, this process is:

1. Initialize centroids $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$.

2. Repeat until convergence:

   - $\forall\ i,\ c_i = \arg\min_j \parallel x_i - u_j \parallel^2$
   - $\forall\ j,\ x_j = \frac{\sum_{i=1}^{m}\{c_i=j\}x_i}{\sum_{i=1}^{m}\{c_i=j\}}$,

where c_i represents each datapoint and $i, j \in \mathbb{R}$. This algorithmic estimation is used to cluster questions together based on the prior results of the respondent. As an individual picks more of a number in a block, they will consequently receive more questions regarding that number in the next block. It takes the individual's response and classifies it for the next block's questions. This survey will place $g = \frac{3}{2}$ as the first centroid with the nearest values becoming the next block's centroid. This is done for the first three blocks, where the final block contains the remaining figures not previously shown.

Figure 2.13 shows the decision tree for the survey after using the k-randomness algorithm, with the numbers representing the g-value, a selected 7 being a solid line, and a selected 1 being a dotted line.

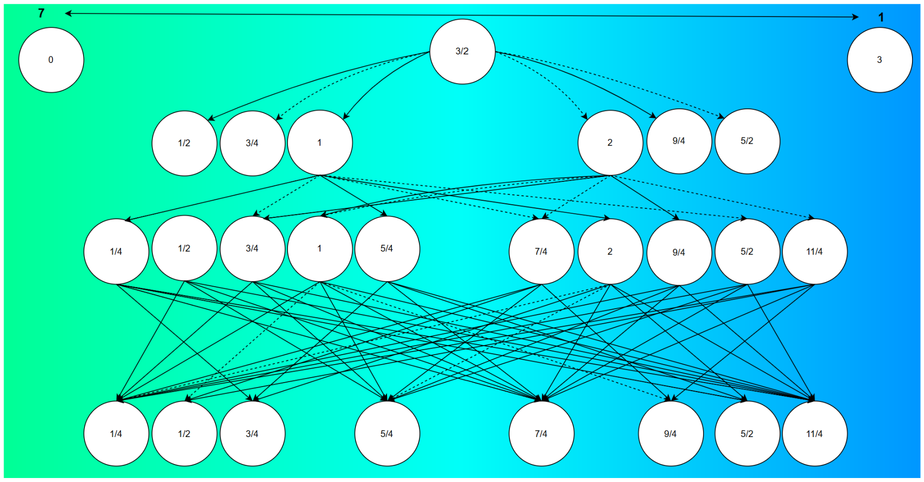

Figure 2.13: The decision tree of the survey with a solid line representing a chosen 7 and a dotted line representing a chosen 1 in front of the HSL green-blue scale.

2.4 Analysis

Now that we have properly created an unbiased survey, we will need to configure the results and understand how to properly assess them. To do this, we will find the *threshold distribution* of the survey.

The threshold distribution views the results at each point of the scale and finds how often each response was selected. The threshold is determined when the response distribution is equal for all variables. For *ismy.blue*, it looked at the distribution of blues. Figure 2.14 shows how, as the scale increases, the frequency of individuals selecting blue increases as well.

Since the results are dichotomous, the responses have to be split in half in order for the threshold to be adequate. The HSL scale showed a threshold distribution

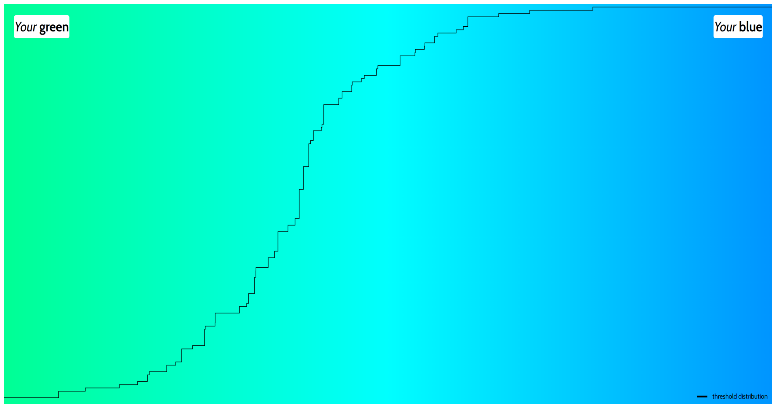

Figure 2.14: Threshold distribution of responses from *ismy.blue* with the HSL scale.

of 175. Based on the aforementioned scale of $[150, 210]$, this number is slightly on the greener side. Thus, individuals are more likely to say that the middle value, 180 is blue rather than green. Similarly, the boundary of people's perspectives between these two colors is slightly more green than the actual boundary.

By finding the threshold distribution from the responses of our survey, we can determine whether individuals boundary between 7 and 1 is different than our created boundary. With this, we can determine if individuals are more likely to state that the middle figure, where $g = \frac{3}{2}$, is 7 or 1. This will be done by looking the the frequencies of 7's.

CHAPTER 3

RESULTS

The survey received 225 responses over the span of three weeks. Of these 225, 202 were sufficiently completed. These sufficient responses answered all 26 questions.

We will begin by viewing the first section of questions. Below is the table of the frequencies of a 7 being selected at each value of g in the first section of questions.

Value of g	7 Frequency	7 Percentage
0	201	99.505%
$\frac{1}{4}$	189	93.564%
$\frac{1}{2}$	177	87.624%
$\frac{3}{4}$	120	59.406%
1	137	67.822%
$\frac{5}{4}$	108	53.465%
$\frac{3}{2}$	25	12.376%
$\frac{7}{4}$	50	24.752%
2	19	9.406%
$\frac{9}{4}$	11	5.446%
$\frac{5}{2}$	9	4.455%
$\frac{11}{4}$	8	3.960%
3	0	0.00%0

From Table 3, it appears that the value of g where the distribution of 1 and 7 are equal is slightly larger than $g = \frac{5}{4}$, as the percentage of 7's is slightly above 50%

and drastically falls below that margin at $g = \frac{3}{2}$. Figure 3.1 shows the threshold distribution of 7's compared to the figures at each value of g.

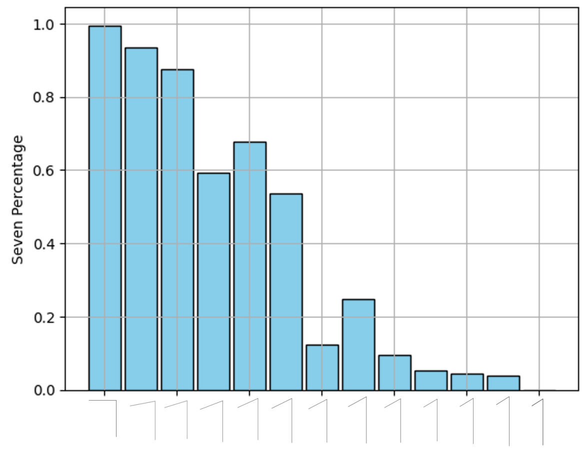

Figure 3.1: Graph of the frequencies of a 7 being selected at each pictured figure.

The threshold occurs slightly left of the middle value $g = \frac{3}{2}$. Thus, for the first section, individuals are more likely to state that the figure at $g = \frac{3}{2}$ is a 1 rather than a 7. Similarly, the boundary of people's perspectives between these numbers is slightly framed towards 7 than the true boundary between these two.

Next, let us look at the frequencies of 7 in the second section, where an additional horizontal line is placed. Below is the table of frequencies of 7 being selected at each value of g in the second section.

Value of g	7 Frequency	7 Percentage
0	201	99.505%
$\frac{1}{4}$	202	100.00%
$\frac{1}{2}$	201	99.505%
$\frac{3}{4}$	199	98.515%
1	195	96.535%
$\frac{5}{4}$	189	93.564%
$\frac{3}{2}$	179	88.614%
$\frac{7}{4}$	150	74.257%
2	106	52.475%
$\frac{9}{4}$	85	42.079%
$\frac{5}{2}$	19	9.406%
$\frac{11}{4}$	1	0.495%
3	0	0.000%

From the table above, the value of g where the distribution of 1 and 7 equal is slightly above 2, as the percentage of 7's is slightly above 50% and falls below the margin at $g = \frac{9}{4}$. Figure 3.2 shows the distribution of 7's compared to the figures at each value of g.

The threshold occurs to the right of the middle value $g = \frac{3}{2}$. Thus, for the second section, individuals are more likely to state that the figure at $g = \frac{3}{2}$ is a 7 rather than a 1. Similarly, the boundary of people's perspectives between these numbers is heavily framed in the direction of 1 compared to the true boundary.

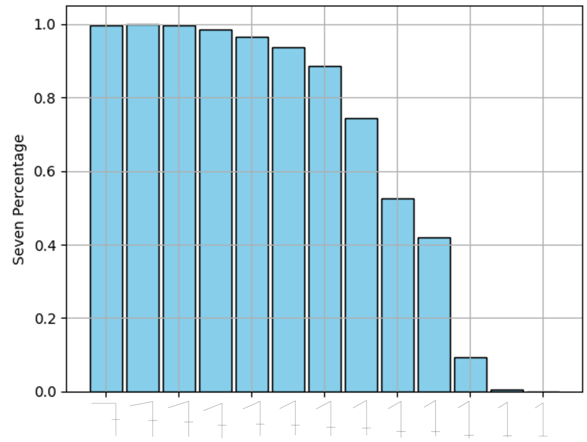

Figure 3.2: Graph of the frequencies of a 7 being selected at each pictured figure in the second section.

For the second section, the distribution of 7's at each value of g were much higher compared to the first section. This may be due to the connection of the horizontal line and 7. More individuals write a horizontal line for 7's rather than 1's. Thus, it is understandable that the horizontal line is connected with 7 rather than 1.

CHAPTER 4

CONCLUSION

This research determines the threshold between two numbers, 7 and 1, through a scaling system based on the magnitude of g determined by matrix multiplication. The threshold was determined by responses from an unbiased survey created through a k-randomness algorithm. These responses were analyzed through a threshold distribution, where it was determined at what point in the scale the threshold occurred. In the first section, we can conclude that the threshold between the numbers 7 and 1 occurs slightly above the value of $g = \frac{5}{4}$, which means that the boundary for most individuals shifted towards the figure at $g = \frac{3}{2}$ being 1 rather than 7. In the second section, we can conclude that the threshold is slightly above the value of $g = 2$, which means that the figure at the middle value of $g = \frac{3}{2}$ is a 7 rather than a 1.

APPENDIX: Matrix Multiplication

Here is the matrix multiplication for the initial vertical shearing transformation.

$$\begin{bmatrix} 1 & 0 \\ g & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \cdot x_1 + 0 \cdot y_1 \\ g \cdot x_1 + 1 \cdot y_1 \end{bmatrix} = \begin{bmatrix} x_1 + 0 \\ g x_1 + y_1 \end{bmatrix} = \begin{bmatrix} x_1 \\ g x_1 + y_1 \end{bmatrix}$$

Here is the matrix multiplication for the displacement and shearing matrix on the vector.

$$\begin{aligned}
\begin{bmatrix} 1 & 0 \\ 0 & \frac{y}{gx_1+y_1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ g & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0 & \frac{y}{gx_1+y_1} \end{bmatrix} \begin{bmatrix} 1 \cdot x + 0 \cdot y \\ g \cdot x + 1 \cdot y \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & \frac{y}{gx_1+y_1} \end{bmatrix} \begin{bmatrix} x \\ gx + y \end{bmatrix} \\
&= \begin{bmatrix} 1 \cdot x + 0 \cdot (gx+y) \\ 0 \cdot x + \frac{y}{gx_1+y_1} \cdot (gx+y) \end{bmatrix} \\
&= \begin{bmatrix} x + 0 \\ 0 + \frac{y}{gx_1+y_1}(gx+y) \end{bmatrix} \\
&= \begin{bmatrix} x \\ \frac{y}{gx_1+y_1}(gx+y) \end{bmatrix}
\end{aligned}$$

REFERENCES

[1] D. Austin. *Understanding Linear Algebra*. 619 Wreath, 2023.

[2] B. Hillmer. Randomize questions. 2024. *https://help.alchemer.com/help/randomize-questions*.

[3] A. Kaur. Shearing in 2d graphics. 2024. *https://www.geeksforgeeks.org/shearing-in-2d-graphics/*.

[4] P. Mineault. Is my blue your blue? *https://ismy.blue/*.

[5] J. Moszczynski. The multi-individuality of handwriting. *Forensic Science International*, 294:e4–e10, 2019.

[6] C. Piech. K means. 2013. *https://stanford.edu/cpiech/cs221/handouts/kmeans.html*.