Distance Estimation:

Exploring Gestalt Principles when Engaging in Tree Thinking

Shannon Hurlston

Vanderbilt University

March 15, 2019

Abstract

Gestalt principles of grouping have been widely studied in diagrammatic reasoning because diagrams that follow these principles are easier to understand than those that do not. In the present study, I investigate how Gestalt principles of grouping can be applied to hierarchical diagrams. Participants were shown an image on a computer screen of a cladogram with two arrows pointing to two branches and were asked to estimate the distance between them by drawing a line using the arrow keys. I hypothesized that same group trials, or distances that followed the Gestalt grouping principles, would be estimated as shorter than those that were not aligned with the Gestalt grouping principles or different group trials. The results were inconclusive due to the many significant interactions between the variables studied. Future directions for this study include rewriting the computer program to improve the user experience, increasing the amount of time participants have to encode the structure of the cladogram, and decreasing the size differences between the cladograms.

Impact of Gestalt Grouping on Cognitive Tasks

Gestalt psychological principles have been at the forefront of perception research since the late 1800s. These theories identify strong tendencies for our brain to group objects into recognizable patterns so that it can organize and process images in the visual space. For example, the grouping principle of connectedness states that objects that are connected to one another are perceived as a group whereas unconnected objects are not. Grouping theories also affect the judgements that we make on cognitive tasks. Woodman, Vecera, and Luck (2003) demonstrated that one of the main benefits of these grouping principles is that they allow the brain to store more information in working memory. Landy and Goldstone (2010) showed that grouping principles affect how students answer math problems. They found that individuals used spacing cues aligned with the Gestalt principle of proximity to solve complex math problems rather than solving them by using order of operation. Bolte and Goschke (2006) built off the ideas of Bowers, Regehr, Balthazard, and Parker (1996) to understand how implicit priming would help individuals identify congruent or incongruent images with Gestalt grouping principles. In their experiment, the researchers showed participants an image consisting of lines and shapes for three seconds. The participants decided whether the fragmented image formed a picture, or if it was just a random series of shapes. Bolte and Goschke (2006) discovered that participants identified congruent images more often than incongruent images, even if they could not recognize the image itself. Due to the strong links between Gestalt principles and a variety of cognitive tasks, I believe these principles may also be related to student’s learning.

Importance of Gestalt Principles in Education

Gestalt principles have been shown to benefit students’ learning in a variety of fields. Lemon et al (2007) studied how engineering students interpreted software architecture diagrams.

They tested two simple software architecture diagrams and one complex diagram by asking participants to answer 20 questions about each diagram's function. Their results indicated that the more complex diagrams were understood with greater accuracy when they employed the Gestalt grouping principles of proximity, similarity and continuity. They also found that reaction time to answering the questions was reduced when participants were examining diagrams aligned with Gestalt grouping principles than when they were misaligned.

Stieff (2007) sought to facilitate student's learning of organic chemistry using molecular diagrams. He tested students on their ability to look at rotated images of symmetric and asymmetric molecules to determine whether both images were the same or different. He conducted a series of experiments to determine how students engage in mental rotation tasks in order to answer stoichiometric problems. In this research, he compared the strategies of two groups: expert chemists, defined as individuals with PhDs, and novice students in organic chemistry. He found that experts were more adept at engaging in mental rotations than students, but both groups were consistently faster at mental rotations using symmetrical diagrams than asymmetrical diagrams. This shows that Gestalt principles can aid learning because diagrams that follow the principle of symmetry were easier to mentally rotate leading to more accurate answers on the follow up questions than those that were asymmetrical.

Novick and Catley (2007) discovered the value of Gestalt psychology as it relates to comprehending a type of evolutionary tree diagram called a cladogram. A cladogram shows the evolutionary relatedness between multiple animals or plants, known as taxa (see Figure 1). They found that because of the grouping principle of good continuation, students had a greater understanding of the diagram if it was presented in a rectangular format (as in Figure 1) instead of a diagonal format. Novick has also studied the effects of grouping on the way people respond to questions based on relationships among taxa. For instance, Novick and Fuselier (2018) presented participants with two cladograms that showed the same relationship between three taxa but in two different ways. They asked participants to identify which cladogram showed the relationship between these taxa more accurately, or if both diagrams showed the same relationship. They found that when the two related taxa were put in the "same group," meaning that two taxa are included in one grouping, and the third taxa was put in a different grouping, (see Figure 1a) students were more likely to identify this relationship as "more accurate" because it followed the Gestalt principles of grouping rather than a correct interpretation of the diagram. The purpose of the present research is to answer the fundamental question of why students make conceptual errors when engaging in diagrammatic reasoning using a cladogram. I hope to clarify these errors so that this model serves as a visual aid rather than a confusing hindrance to learning.

Boundary Effects in Segmented Diagrams

Maps are a specific type of diagram that not only use Gestalt grouping principles to segment a space into specific areas such as county lines or city blocks, but these separations also express relative distance between two points. Maki (1981) examined how a map's boundaries affected reaction time. She studied the distances between 12 east/west pairs of cities across two states, North Dakota and Minnesota. Participants were first asked to learn the names and locations of each city before they were tested on a series of true/false relationships about the relative locations between them. She found that when the cities being compared were in different states, the participants' reaction times to the true/false questions about two cities' relative locations were slower than if they were in the same state. She argued this difference is caused by the boundary between the two states serving as a perceptual "barrier" between two cities on opposite sides of the border. This effect was shown as well when the participants learned “fake names” of cities. This research demonstrated that there are strong grouping effects within a diagram. It takes longer to perceive relative location between two cities in different states if there is a barrier separating them, presumably because the cities have been grouped together due to the Gestalt principle of same region.

McNamara (1986) similarly studied how maps can affect distance estimations and directional judgements. He organized various objects in a “spatial layout” that was divided into four distinct regions of test objects and filler objects. He designed a room in which participants would be allowed to freely move about to study the objects, have limited or constrained movement in the space, or given a map in order to memorize the locations of all objects. The participants were subsequently judged on their accuracy in correctly placing the test objects in the room. The participants were significantly faster in placing two objects when the objects were in the same region of the map than if they were in different parts of the map. Their response times were also faster when the objects were closer to rather than further away from each other. When two objects were close in proximity to one another, participants were more likely to think that the distance between them was greater than it actually was. This is due to the phenomenon known as the boundary effect because the two objects were visually separated by a barrier and therefore were perceived as farther apart because they were in different regions. However, participants were less likely to overestimate distance when objects were farther away from each other but in the same region, demonstrating that the closer the objects are, the more “attractive” the boundary effect is. These results strengthens Maki’s (1981) hypothesis that boundaries change how people perceive relative distance because McNamara (1986) confirmed that participants overestimate the distance between two objects separated by a barrier.

Irmak, Naylor, and Bearden (2011) tested how businesses could be affected by boundary effects. Participants were given a starting point on a map and subsequently shown two store locations. One store was closer in terms of absolute distance from the starting point but was outside the city limits. The other store was farther away but was within the city border. They found that although the absolute distance was shorter, participants would prefer to shop at a store within their city limits than a store outside their city limits. In a follow-up study, they found that people were willing to pay more for plane tickets to a destination outside of a region (from the South to the Northeast) even if the absolute distance between the two cities was identical.

McNamara (1986), Irmak et al. (2011), and Maki (1981) all demonstrated the power of boundary effects on the perceptual organization of distance. Our brains naturally categorize objects into belonging to the same or different groups. All three studies showed that items within a boundary were perceived as belonging to the same unit and that participants were quicker and more accurate at identifying them. In contrast, when cities or objects were on two different sides of a boundary, participants separated them into different groups even if the relative distance between the two points was equidistant. When participants overestimate distances in different groups, their overestimations shows that the boundary creates a divide in our perception of the image that is difficult to override. Our tendency to group objects is apparently so automatic that we see two things that are the same distance apart as separated simply because of the division. The present study will examine how or if boundary effects effect the distance estimations of hierarchal diagrams.

Distance Estimation Tasks

Coren and Gurgis (1980, 1985) examined how distance is perceived between two dots that have the exact same distance but differ in the way they are organized in relation to surrounding dots. In their experiment, participants saw a triangle made of several, equally spaced dots and one dot that was not a part of the triangle, but was the same distance apart as the dots belonging to the triangle. They were asked to judge the “inside distance” between two dots following the outline of the triangle and the “outside distance” between the dot outside the pattern and one that belonged to the triangle. The inside distances aligned with grouping principles such as proximity and good continuation. They hypothesized that the distances that did not follow the grouping principles would be perceived as “longer” than the ones that were consistent with the grouping principles, even though the distances between the two points were both 40 mm in length. The distance misaligned with the Gestalt principle was perceived as being 3% longer than the distance consistent with the Gestalt principles. They tested this using the principles of proximity, similarity, closure, and good continuation and found that all four principles had statistically significant results that confirmed their hypothesis. I will attempt to replicate their results as they relate to the rectangular format cladograms.

Our brains often look for the shortest relative distance between two objects in order to organize a pattern in visual space. Kubovy and Wagemans (1995) tested a hypothesis that individuals would perceive the structure of a dot lattice by the closest distance vector between the two dots. The participants saw a circle with a pattern of dots shown on the inside and were asked to select how the dots were organized using a line oriented at different angles to show the direction of the grouping. They discovered that participants reliably selected the orientation of the line which demonstrated the shortest distance vector between dots. This shows that the Gestalt principle of proximity is especially relevant when determining how our brains interpret simple patterns.

Healey, Tack, Schnieder, and Barshi (2015) explored how different types of training would affect participants' accuracy on distance estimation tasks. Participants either received training on how to judge distance based on the time it took for a line to extend from one endpoint to another, or based on the length of the line itself. The participants were then asked to either complete a distance estimation task using the same or opposite strategy employed in the training. The researchers found that participants were more accurate at assessing the distance of a line when the training they received matched the experimental condition. Additionally, they found that "learning" occurred during the training because regardless of the training condition received, participants were able to apply their training to the opposite experimental condition, although their reaction time was slightly slower. This shows that distance estimation, as a crucial element of our study, can be measured both temporally and through length estimations. The present study asks participants to perform the length estimation task as a line stretched between two endpoints defined by the user.

Overview of the Present Study

The goal of my research is to understand why students have difficulty understanding cladograms. Cladograms are an essential tool for learning in the field of Evolutionary Biology, and these diagrams are a key visual aid in demonstrating relationships among taxa. Students have difficulty understanding these diagrams for a variety of reasons that I believe stem from the Gestalt grouping principles. If these diagrams are misinterpreted due to the grouping principles, this will give me some key insight on how to better display this information so that students can better understand the topic. This research will allow me to further explore the effects of Gestalt perceptual principles on the interpretation of cladograms to better understand students' reasoning when engaging in tree thinking or interpreting this diagram.

The main research question that I am going to test through this study is “How are relative distances between taxa perceived on cladograms?” In order to answer this question, participants will estimate distances between two branches on a cladogram that are either aligned with the Gestalt principles of grouping or not. During this experiment, participants will be asked to emulate the Coren and Gurgis (1980) distance estimation task on a computer by drawing a line to reflect how long they think the distance is between two branches indicated by red arrows on a cladogram.

Same group trials, or distances consistent with Gestalt Principles are defined as what biologists refer to as “sister groups”. The distance is enclosed by two branches that are directly connected to each other (see Figure 2, blue). Sister groups look like a three sided square because they consist of two branches that are directly connected to each other. Different group distances are separated by vertical distance and consist of two taxa that are not directly connected to each other (see Figure 2, red and green). My hypothesis is that the horizontal distance between two branch tips will be perceived as longer if the branch tips are part of different groups than if they are part of the same group. This research is extremely important because it expands our current thinking about how students interpret cladograms. If I can show that Gestalt principles are at play in the interpretation of these diagrams, it will provide important insight into the types of mistakes students are making when engaging in diagrammatic reasoning. If students rely on Gestalt perceptual information rather than the diagram’s visual information to engage in tree thinking, this lends itself to misunderstandings about the relationships between the taxa and ultimately shows the gaps in student knowledge when being taught this material.

Method

Participants

Participants were 30 male and female Vanderbilt graduate and undergraduate students (6 graduate, 24 undergraduate; 5 male, 25 female) aged 18-25 recruited from SONA. 15 participants were white, 6 were Asian, 6 were Black, and 3 were other. They were compensated $14 for their time. The data from one graduate student participant were excluded from the analyses because he produced too many outlier responses, so only 29 participants were included in the data analyses.

Apparatus

I used a computer program written in Java that presented the stimuli and drew a line through user input on the right and left arrow keys. This line was measured by the program in pixels and saved to a .txt data file. The data collection was performed on a 21.5 in Mac desktop computer running OS 11.10 using an Intelij compiler community version 2018.3.

Stimuli and Design

I used four six branch trees, four seven branch trees and four eight branch trees (see Appendix A) as test stimuli for this study and four five branch trees as stimuli for the practice trials. The 12 test trees were presented in three distinct sizes with 10, 15, and 25mm separating each branch. I chose to vary the distance between the branches so that participants did not learn to estimate the same distance but rather provided a unique estimation every time. All variables were manipulated within subjects.

The primary independent variable was whether subjects were asked to estimate the distance between two branches that were in the same group or different groups. There were 72 same group trials (see Figure 2, blue) and 72 different group trials (see Figure 2, green and red).

Same group trials look like a three sided square in which the two branches are connected by a line at the bottom. In evolutionary biology, the two smaller branches are often referred to as a *sister group* because the two taxa share the same most recent common ancestor. For the same group trials, 36 were on the outer edge of the cladogram (see Figure 6, green) which classified them as *exterior trials* (SE) and the other 36 were *interior trials* (SI) located on two branches in the interior of the cladogram (see Figure 6, blue).

Different group trials consisted of two branches that are not directly connected to each other (see Figure 2, green and red). There were two types of different group trials. Sister group/singleton were trials in which participants estimated the distance between one branch of a sister group and a single branch (see Figure 2, green). Singleton/singleton were trials in which the arrows pointed to two single branches (see Figure 2, red). Each of these trial types occurred equally often on the interior and the exterior of the cladograms. Thus, of the 72 total different group trials, there were sister/singleton trials on the exterior (DGRE trials), singleton/sister on the interior (DGRI trials), singleton/singleton trials on the exterior (DGGE trials) and singleton/singleton trials on the interior (DGGI trials). Each of the 12 test cladograms was designed to include two same group trials, one SE and one SI as well as two of each of the four different group trials DGGE, DGGI, DGRE, and DGRI.

Two of the four trees with six, seven, and eight branches had a majority of their branches on the right, and the other two were more heavily weighted on the left (see Figure 3). The weight of a tree is the direction the cladogram would “fall” if placed on a balance scale. Finally, two trees at each branch size were classified as either having two groups or not two groups. A tree classified as having “two groups” showed a clear division between the branches in which the diagram is separated into two distinct parts (see Figure 4a). “Not two group” trees do not show a clear distinction (see Figure 4b).

I also varied the trials so that each of the six trial types had an equal number of arrows pointing to the branches positioned on the center, right, or left side of the cladogram. The different group trials differed in the amount of “vertical distance” between each of the two branches. I considered vertical distance in three ways (see Figure 5). The first measure of vertical distance was the number of “steps” from one branch to another (see Figure 5.1). The greater number of steps from one branch to another, the larger the vertical distance. The second measurement was derived by taking length of the longest branch in the pair (see Figure 5.2). The longer the branch was, the larger the vertical distance became. Finally we observed the varied amount of “white space” or area in square millimeters between the two branches (see Figure 5.3). The greater the area of white space, the larger the vertical distance was.

The four five branch practice trees were shown with 12.5 and 20mm between each branch. The practice trials were similarly divided into four same and four different group trials. There were 16 practice trials total.

Trial Blocks

The 144 trials were divided into four blocks of 36 trials. Trials were randomly assigned to blocks subject to the following constraints. Eighteen trials in each block had to be “same group” and 18 trials had to be “different group”. There were roughly equal numbers of SI and SE trials in each block (defined as no fewer than eight of either trial type). Similarly, nine different group trials were interior trials, and nine were exterior trials. Among the nine trials in the same location on the cladogram, at least three of each type were presented in a block. Twelve cladograms of each physical size (10, 15, and 20 mm) were in each block. The number of branches were assigned to blocks roughly equally so that no more than 13 trials in each block had the same number of branches (6, 7, and 8). The weighting of the trials was assigned roughly equally so that no more than 20 trials had the same weight (left/right).

The participants received the trials in one of two orders. Order 1 presented blocks 1, 2, 3, 4 in order from trials 1-36 in each block. Order 2 presented blocks 3, 4, 2, 1 in order from trials 36-1 in each block. Half of the participants were randomly assigned to each order.

The order of the trials within each block was random subject to the following constraints. No two of the same trial type (DGRI, DGRE, DGGE, DGGI, SE, or SI) appeared back to back. Approximately half of each trial type appeared in the first 18 trials and the other half in the following 18 trials. For size, weighting, grouping, and number of branches, no more than three of the same category could be presented back to back.

Procedure

After participants signed a consent form, the experimenter started the appropriate program based on the randomly assigned order. The program displayed the instructions on the screen. After the participant read the instructions on-screen, the experimenter asked the participant if he/she had any questions. The participant completed a block of 16 practice trials using the four five-branch cladograms. Once the practice trials were completed, the program asked participants to continue on to the experimental trials. After completing a block of 36 experimental trials, participants were forced to take a minimum 1 min break in which the program did not permit the user to move on with the study until 1 min had expired. This procedure continued for each of the four experimental blocks. Participants were compensated for their time.

During each trial, a fixation point was presented in the middle of the screen for 500 msec. Immediately following the fixation point, participants were shown an image of a cladogram with two arrows pointing to two adjacent branches for three seconds. The window itself filled the screen so that the participant could not see anything except the cladogram image. The arrows indicated the distance the participant was asked to estimate (Figure 7). After three seconds expired, the program displayed a second full size window. A small black circle appeared in the window at 200 x 200 pixels measured from the top left of screen indicating the starting position of the line. The drawing program allowed participants to estimate the horizontal distance between the branches indicated by the two arrows they saw on the cladogram using the right arrow key to expand, and the left arrow key to contract the line. The line was positioned above where the cladogram was shown so that participants would have difficulty preserving the image in their mind to make the estimation (Sperling, 1960). Participants pressed the red X button in the top left corner of the screen to close the window when they were satisfied with their answer. Participants had 12 seconds to complete each distance estimate and close the window. The program automatically continued after 12 seconds had expired. After they had completed the perceptual portion of the experiment, participants were asked to fill out a demographic survey on REDCap which asked about age, race, year in school, major, and biology courses taken. The study took approximately 50 minutes to complete.

Results

Before analyzing the results I engaged in a data cleaning process. I divided the trials into groups by size and found the outlier estimates using 3 standard deviations from the mean for the cut off for each size. In total, I deleted 42 outliers: 17 from size 10, 21 from size 15, and 4 from size 25.

After completing the data cleaning process, I ran a descriptive statistics analysis on the raw distance estimates and found that the variances for the three cladogram sizes were very discrepant: 55.25, 102.43, and 249.21 respectively. Additionally, participants on average underestimated the distances for each trial size. 10mm converts to 37.795 pixels, 15mm is equivalent to 56.693 pixels, and 25mm is equivalent to 94.488 pixels. The results will be reported in pixels because the computer program measured the distance estimations in this way. On average participants estimated the distances between adjacent branches on 10mm cladograms to be 36.49 pixels or 96% of the original length, 15mm to be 51.24 pixels or 90.38% of the original length, and 25mm to be 79.46 or 79.46% of the original length. I decided to transform the length estimations using the natural logarithm in order to better equate the variances. I used these transformed scores in conducting the ANOVA tests.

Initial Analyses

I first ran a 3x2x2x2 within-subjects ANOVA to analyze the differences between size (10 vs. 15 vs. 25mm), location (interior vs. exterior), grouping (two groups vs. not two groups), and trial type (same vs. different group) in order to test whether different group trials lead to longer distance estimations than same group trials. As shown in the ANOVA table in Appendix B, all the main effects and interactions, including the four-way interaction, were significant except for the main effect of grouping. In an attempt to provide some more descriptive information about participants' distance estimations, I conducted separate ANOVAs on the same and different group trials as well as by size.

For the analyses by size, I ran three 2x2x2 within-subjects ANOVA to compare trial type, grouping, and location for each size cladogram. I found that with increasing size, there were fewer significant results (see Tables 2, 3, and 4). For size 10, all main effects and interactions were significant (see Table 2). For size 15, there were main effects of grouping and trial type as well as significant interactions between grouping and location and grouping, location, and trial type (see Table 3). For size 25, only the interaction between grouping and trial type was significant (see Table 4).

When thinking about the best way to analyze the data given all of the significant results, I referred back to the hypothesis that same and different group trials would have differing distance estimations due to the Gestalt principles of grouping. Ultimately, I am not interested in how size affects the distance estimations that participants gave because it is fairly obvious that the larger the distance between the two branches, the larger \ the participant would estimate the distance to be. Therefore, the follow up analyses that made the most sense to focus on are the separate ANOVAs for same-group trials and different-groups trials.

Same Group Analysis

I conducted a 3x2x2 within-subjects ANOVA comparing size, location, and grouping. I found that the same group trials were the causes for most of the significant results in the primary analysis. Table 6 displays the results for the same group trials.

I found that all the main effects and interactions were significant for the same group trials. The main effect of size indicated that larger sized cladograms received bigger distance estimates. Two-group trials (*M = 57.36*) were measured as longer than the not-two-group trials (*M = 55.33*). The interior trials (*M = 55.28*) were estimated as shorter than the exterior trials (*M = 55.72)*.

There was an interaction between size and grouping in which for size 10 (*M = 35.92, M = 36.58*) and size 15 (*M = 51.15, M = 52.44*) the two group trials were estimated as larger than the not two trials. In contrast, for the 25mm cladograms, the not two groups (*M = 76.94*) were estimated as longer than the two groups (*M = 76.08*) (see Figure 8). There was an interaction between size and location in which there was no difference between the size 10 interior (*M =*

36.44) and exterior trials (*M=36.07*) and the size 25 interior (*M = 76.53*) and exterior trials (*M = 76.50*). However with size 15, exterior (*M = 51.53*) was measured as shorter than interior (*M = 57.69*) (see Figure 9). There was an interaction between grouping and location in which there was no difference between the not two group trials on the interior (*M = 54.60*) and exterior (*M = 54.74*), but for the two group trials the interior trials (*M = 55.96*) were measured as longer than the exterior trials *(M = 54.92*) (see Figure 10).

Finally, there was a three-way interaction between size, grouping, and location in which there was no difference between size and grouping and location for size 10 (see Figure 11). There was however, a significant difference between size and grouping for size 15 and size 25. For size 15, the interior trials were estimated as slightly smaller than the exterior trials for both the two group (*M= 47.52, M=51.21*) and not two group trials (*M = 50.33, M=51.96*) (see Figure 12). For size 25, the opposite was true as the exterior trials (*M = 76.16*) was estimated as longer than the interior trials (*M = 66.05*) on not two group cladograms (see Figure 13).

Analysis of the Different Group Trials

For the different group trials, I analyzed the data using a 3x2x2x2 within-subjects ANOVA looking at size (10 vs. 15 vs. 25 mm), location (interior vs. exterior), grouping (two groups vs. not two groups), and trial type (singleton/singleton vs. singleton/sister). There was a significant main effect of size (see Table 5). Obviously, size 10 trials were estimated as shortest (*M = 35.92*), followed by size 15 (*M = 49.66*), and size 25 had the largest mean (*M = 75.33*). There was also a main effect of grouping where two group trials (*M = 54.26*) were estimated as longer than not two group trails (*M = 53.02*). Finally, there was a main effect of trial type. Singleton/singleton trials (*M = 54.21*) were estimated as longer on average than singleton/sister trials (*M = 53.07*). There were no significant interaction effects.

Discussion

Our hypothesis was that same group distances would be estimated as shorter than different group differences, consistent with the results of Coren and Gurgis (1985) because these distances followed the Gestalt grouping principles. The hypothesis was not supported. Given all of the significant two, three, and four-way interactions, a consistent interpretation of the results seems impossible.

When I conducted the separate within-subjects ANOVAs for same and different group trials, I found that most of the significant differences in the results came from the same group trials rather than the different group trials. This is interesting because different group trials have two subsets whereas same group trails have only one. When I examined the means of all the different trail types, I found that same group trails were estimated on average as 54.94 pixels, singleton/singleton were estimated on average as 54.21 pixels, and singleton/sister were estimated as 53.07 pixels on average. This result is opposite of our hypothesis that same group trials would be estimated as shorter than different group trails because the same group trials are consistent with the Gestalt principles of grouping. However, this result is hard to interpret because when broken out by size, the results flip depending on the size of the cladogram.

Participants generally underestimated the distance between the two arrows. On the smaller cladograms, this estimate was more accurate as the 10mm cladograms were underestimated only by about 4% whereas the larger cladograms were underestimated by around 21% This is opposite of the finding in the Coren and Gurgis (1985) paper in which participants overestimated the 40mm distance between the two dots for both the inside and outside distances. A possible explanation for this is that the 12 second timer caused participants to feel more "time pressured" when the cladograms were larger because it takes more time to draw a longer line.

This would explain why participants underestimated the 25mm cladograms more than the 10 or 15mm cladograms. In order to address this problem, the data collection program would have to be reprogrammed in order to remove this time limit so that participants could take as much or as little time as they needed to complete the task.

The first statistical analysis that I conducted, because of all the significant main effects and interactions that it yielded, showed me that the data would need to be broken out by specific factors in order to gain a more meaningful understanding of the results. Although I varied much more than size, trial type, location, and grouping when creating the stimuli, in order to analyze the results with enough power, I decided to ignore weighting and number of branches because they were not factors that I felt would significantly alter the way that participants estimated the distance between the branches but rather the visual appearance of the cladogram itself.

By conducting the within-subjects ANOVAs for each cladogram size, I learned that as the cladograms got larger, there were fewer significant results. Size 10 yielded the most significant results, followed by size 15 and then size 25. This could be due to a variety of reasons. One possible interpretation is that there is something unique about smaller cladograms that causes participants to systematically vary how they respond to each trial depending on its location, grouping, and trial type. There could be an “attractive force” that pulls the branches together on the smaller cladograms that is not seen on the larger cladograms. As they get larger, this force becomes weaker, and the branches no longer feel like they are being pulled together or apart by the location of the two branches or grouping of the cladogram.

Future Directions

In the future, I would do several things differently in order to better test this hypothesis. First, I would rewrite the program in order to make it easier for participants to use. Currently, users have to close the distance estimation window in the top left-hand corner after finishing every trial. This is difficult because it is not an intuitive way for the participants to move on to the next trial and requires several seconds to move the mouse to the corner and click. Instead, I would like for the participants to be able to press return in order to close the windows because it is more typical of an experiment of this type. Additionally, I would also like to remove the time limit so that participants can complete the task at their own pace instead of not being able to move on until the time expires.

One reason why the results might have been so difficult to interpret is that the participants did not have enough time to encode the overall cladogram structure before encoding which branches to estimate the distances between. In the current study, participants had three seconds to encode the structure and branches of interest at the same time. Because of this, participants might have ignored the structure itself, which is important to understanding how participants think about the differences between the branches. In a future study, I would include an image of the cladogram without any arrows for 2 sec followed by the same image with arrows on it for 1 sec. As the program stands right now, after seeing the fixation point for 500 msec, the participants then see the cladogram with arrows for 3 seconds. By using the same amount of time but breaking it up into smaller parts, we will be able to see how their estimates of distance are affected by differences in cladogram structure more clearly.

Finally, I would vary the sizes of the cladograms less so that the size differences were not so drastic between the three cladograms. The experiment currently uses cladograms of 10, 15 and 25 mm between the branches. In other words, the 15 and 25 mm cladograms are 1.5 and 2 times larger than the smallest 10 mm cladogram. As we saw in the results, the larger the cladograms became, the fewer statistically significant results there were. In a future study, it might be interesting to create cladograms of 1.25 and 1.5x size differences from the original cladogram. For example, if I wanted to test the hypothesis that the smaller cladograms branches are more attracted to each other, I might use 10, 12, and 15 size cladograms. On the other hand, if I wanted to understand which factors are most likely to predict distance estimations in participants I might choose to work with larger cladograms such as 20, 22, and 25mm because they yielded fewer statistically significant results.

References

Bolte, A., & Goschke, T. (2008). Intuition in the context of object perception: Intuitive gestalt judgments rest on the unconscious activation of semantic representations. *Cognition, 108*, 608-616. doi:10.1016/j.cognition.2008.05.001

Bowers, K. S., Regehr, G., Balthazard, C., & Parker, K. (1990). Intuition in the context of discovery. *Cognitive Psychology, 22*(1), 72-110. doi:10.1016/0010-0285(90)90004-n

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27*(3), 183-228. doi:10.3758/bf03204258

Coren, S., & Girgus, J. S. (1980). Principles of perceptual organization and spatial distortion: The Gestalt illusions. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 404-412.

Cromley, J., Weisberg, S., Dai, T., Newcombe, N., Schunn, C., Massey, C., & Merlino, F. (2016). Improving middle school science learning using diagrammatic reasoning. *Science Education*, *100*, 1184–1213. https://doi.org/10.1002/sce.21241

Enns, J. T., & Girgus, J. S. (1985). Perceptual grouping and spatial distortion: A developmental study. *Developmental Psychology, 21,* 241-246.

Grassi, P., Zaretskaya, N., & Bartels, A. (2016). Parietal cortex mediates perceptual Gestalt grouping independent of stimulus size. *NeuroImage*, *133*, 367–377. https://doi.org/10.1016/j.neuroimage.2016.03.008

Hardy, I., Schneider, M., Jonen, A., Stern, E., & Möller, K. (2005). Fostering diagrammatic reasoning in science education. *Swiss Journal of Psychology/Schweizerische Zeitschrift Für Psychologie/Revue Suisse De Psychologie, 64*, 207-217. doi:http://dx.doi.org.proxy.library.vanderbilt.edu/10.1024/1421-0185.64.3.207

Healy, A. F., Tack, L. A., Schneider, V. I., & Barshi, I. (2015). Training specificity and transfer in time and distance estimation. *Memory & Cognition, 43*(5), 736-747. doi:10.3758/s13421-015-0503-9

Irmak, C., Naylor, R., & Bearden, W. (2011). The out-of-region bias: Distance estimations based on geographic category membership. *Marketing Letters*, *22*(2), 181–196. doi:10.1007/s11002-010-9120-3

Kubovy, M., & Wagemans, J. (1998). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, *6*(4), 225–234. doi:10.1111/j.1467-9280.1995.tb00597.x

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *Journal of Experimental Psychology: Human Perception & Performance, 63*(10)

Lemon, K., Allen, E., Carver, J., & Bradshaw, G. (2007). An empirical study of the effects of gestalt principles on diagram understandability. *Empirical Software Engineering and Measurement, 2007. ESEM 2007. First International Symposium on* (pp. 156–165). IEEE. doi:10.1109/ESEM.2007.37

Maki, R. H. (1981). Categorization and distance effects with spatial linear orders. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 15-32.

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology, 18,* 87-121.

Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. *Journal of Experimental Psychology: Applied, 13*(4), 197-223.

Novick, L. R., Fuselier L.C., (2018). Perception and conception in understanding evolutionary trees. Submitted for review.

Wagemans, J., Elder, J. H., Kubovy, M. K., Palmer, S. E., Peterson, M. A., Singh, M., & Von der Heydt, R. (2012). A century of gestalt psychology in visual perception I. perceptual grouping and figure-ground organization. *Psychonomic Bulletin & Review, 138*(6), 1172-1217.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1-29.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction, 17*(2), 219-234. http://dx.doi.org/10.1016/j.learninstruc.2007.01.012

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. Psychonomic Bulletin & Review, 10(1), 80-87. doi:10.3758/bf03196470

Defining Stimuli Manipulations

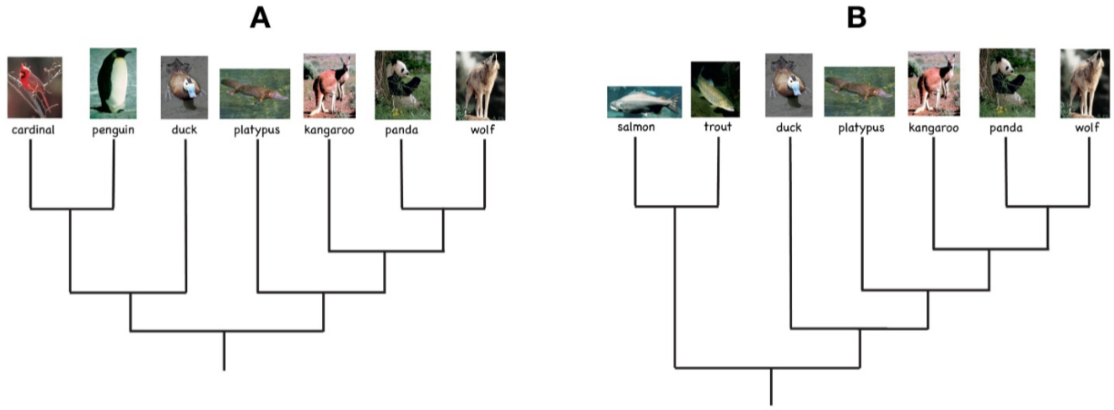


Figure 1: Understanding student difficulty in diagrammatic reasoning from Novick (2018). Students are likely to believe that figure A displays the information that the platypus is more closely related to the kangaroo than the duck because kangaroo and platypus belong to the same group and duck belongs to a different group.

Figure 2: *Green:* different group sister/singleton connection *Red*: different group singleton/singleton connection *Blue*: same group connection

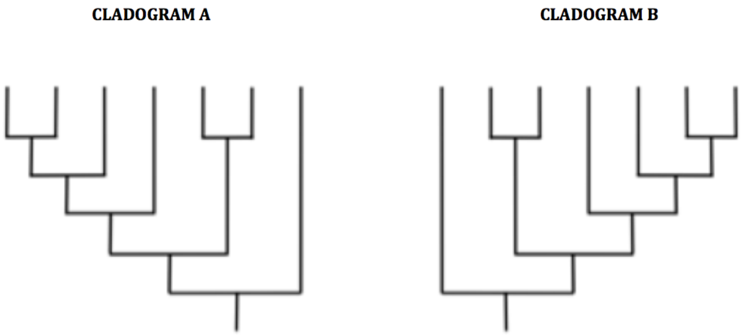


Figure 3: *Cladogram A:* left weighted *Cladogram B*: right weighted

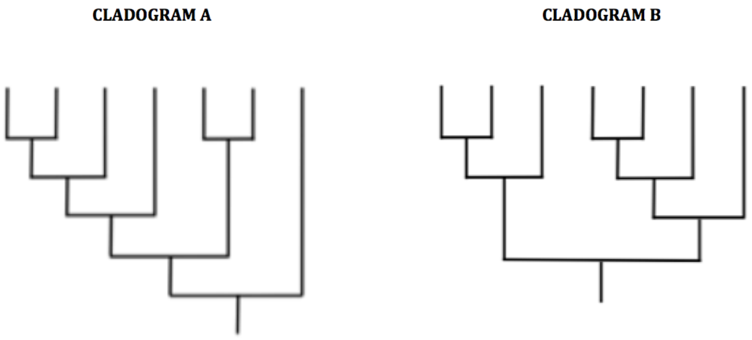


Figure 4: *Cladogram A*: split into multiple parts *Cladogram B*: split into two parts

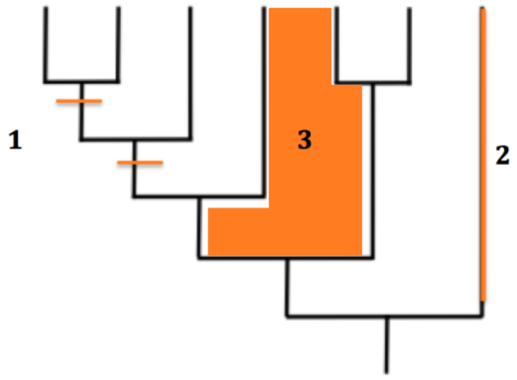


Figure 5: *5.1* number of steps 5.2 vertical distance of longest branch 5.3 area of white space

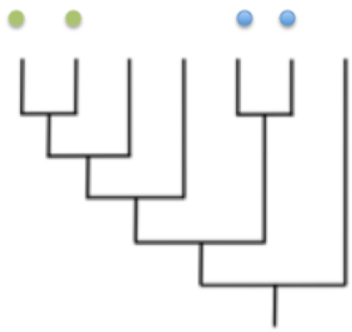


Figure 6: *Green*: exterior trial *Blue*: interior trial

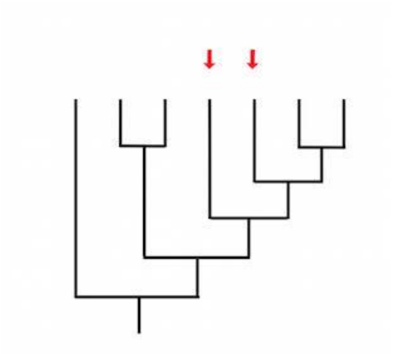


Figure 7: Participants would be asked to draw a line to estimate the distance between the two arrows

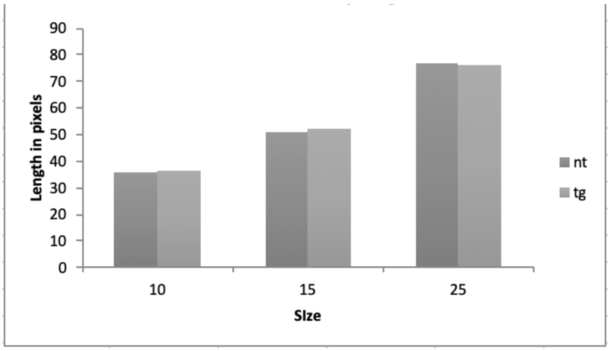


Figure 8: Size x Grouping interaction for all same group trials

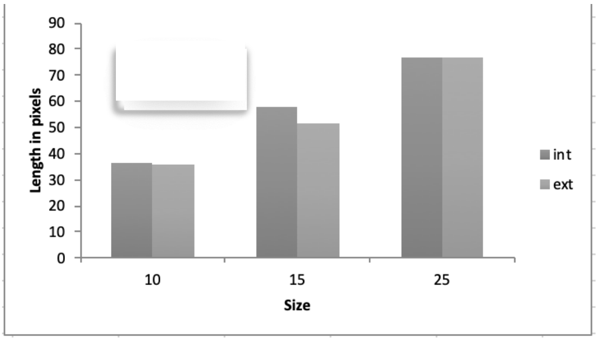


Figure 9: Size x Location interaction for all same group trials

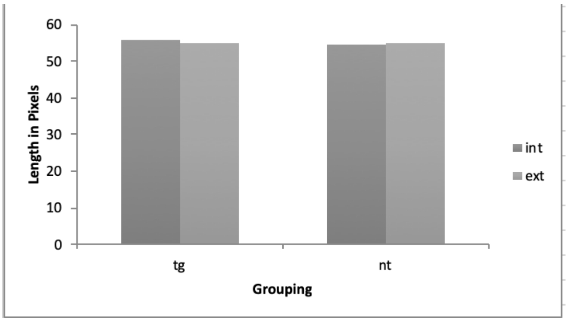


Figure 10: Grouping x Location interaction for all same group trials

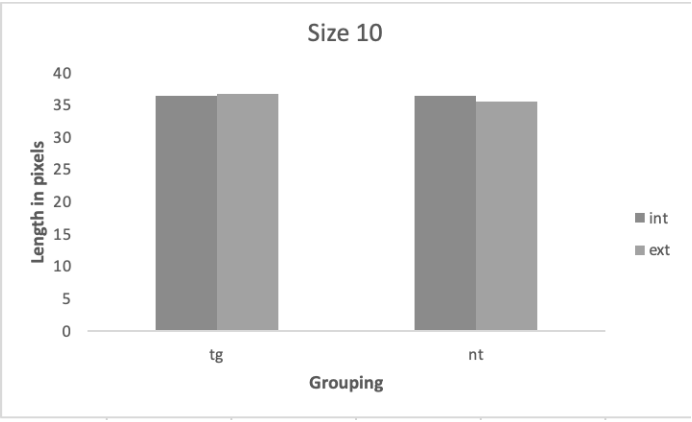


Figure 11: Size 10 x Grouping x Location interaction for same and different group trials

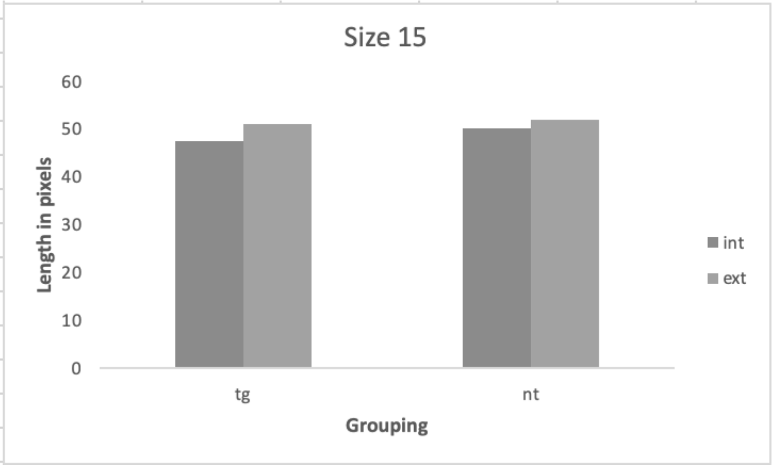


Figure 12: Size 15 x Grouping x Location interaction for same and different group trials

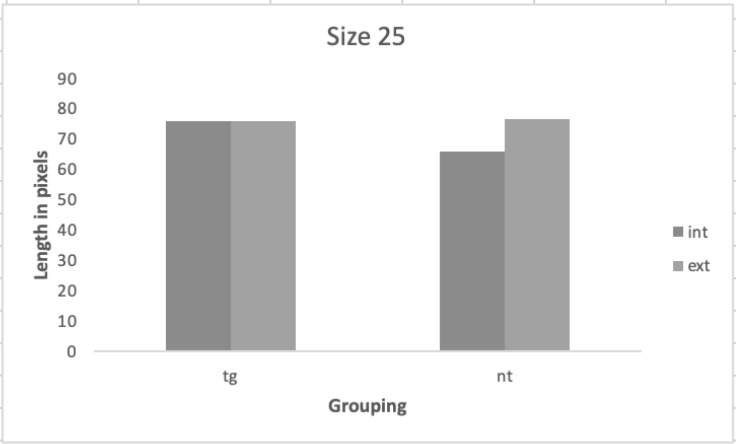
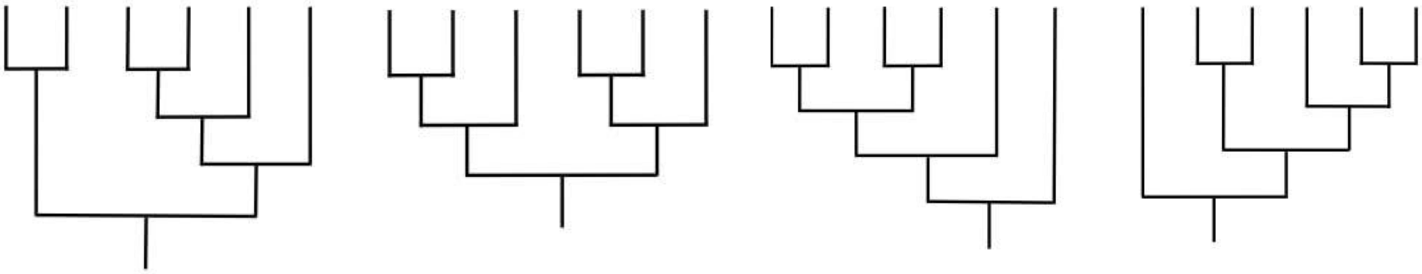


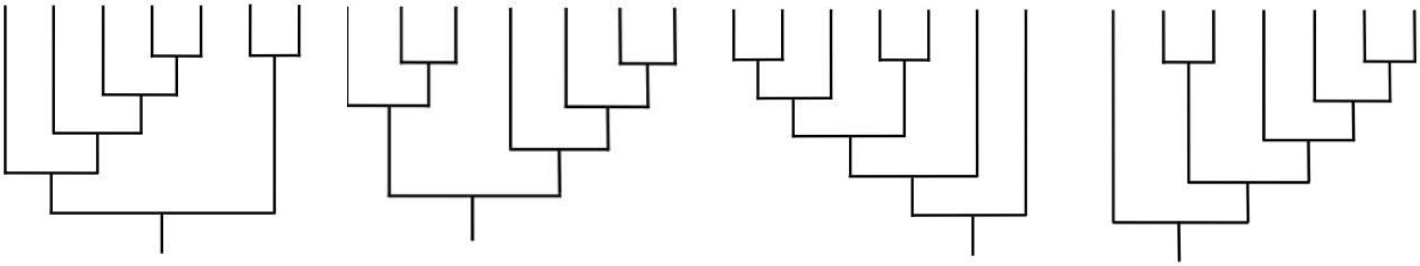
Figure 13: Size 25 x Grouping x Location interaction for same and different group trials

Appendix A
Test Stimuli

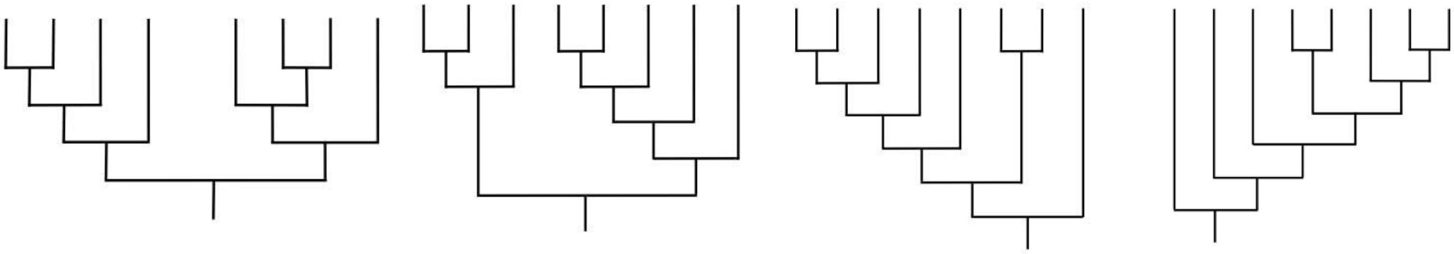
Six Branches



Seven Branches



Eight Branches



Appendix B

Anova Tables

Table 1: ANOVA for Same/Different Group Trials

Source	*SS*	*df*	*MS*	*F*	*Sig*	*Partial Eta^2*
Size	66.446	2	33.223	689.037	.000*	.961
Error (Size)	2.700	56	.048			
Grouping	.000	1	.000	.131	.720	.005
Error Grouping)	.073	28	.003			
Location	.066	1	.066	20.212	.000	.419
Error (Location)	.091	28	.003			
Trial Type	.022	1	.022	5.894	.022	.174
Error (Trial Type)	.104	28	.004			
Size * Grouping	.059	2	.030	9.726	.000	.258
Error (S * G)	.170	56	.003			
Size * Location	.085	2	.042	14.172	.000	.336
Error (S * L)	.167	56	.003			
Grouping * Location	.036	1	.036	13.795	.001	.330
Error (G * L)	.072	28	.003			
Size * Grouping * Location	.036	2	.018	7.436	.001	.210
Error (S * G * L)	.134	56	.002			

Size * Trial Type	.144	2	.072	17.314	.000	.382
Error (S * T)	.232	56	.004			
Grouping * Trial Type	.053	1	.053	17.113	.000	.379
Error (G * T)	.087	28	.003			
Size * Grouping * Trial Type	.058	2	.029	7.254	.002	.206
Error (S * G * T)	.222	56	.004			
Location * Trial Type	.039	1	.039	19.867	.000	.415
Error (Location * T)	.054	28	.002			
Size * Location * Trial Type	.059	2	.030	8.910	.000	.241
Error (S * L * T)	.186	56	.003			
Grouping * Location * Trial Type	.043	1	.043	12.267	.002	.305
Error (G * L * T)	.099	28	.044			
Size * Grouping * Location * Trial Type	.024	2	.012	3.335	.043	.106
Error (S * G * L * T)	.202	56	.004			

---

* Corrected using the Huynh-Feldt correction due to the violation of the sphericity assumption

Table 2: Anova for Same/Different Group Trials at Size 10

Source	*SS*	*df*	*MS*	*F*	*Sig*	*Partial Eta^2*
Grouping	.034	1	.034	14.307	.001	.338
Error (Grouping)	.067	28	.002			
Location	.149	1	.149	48.103	.000	.632
Error (Location)	.087	28	.003			
Trial Type	.037	1	.037	9.777	.004	.259
Error (Trial Type)	.106	28	.004			
Grouping * Location	.040	1	.040	15.624	.000	.358
Error (G * L)	.071	28	.003			
Grouping * Trial Type	.084	1	.084	20.484	.000	.422
Error (G * T)	.115	28	.004			
Location * Trial Type	.097	1	.097	48.940	.000	.636
Error (L * T)	.055	28	.002			
Grouping * Location * Trial Type	.045	1	.045	12.387	.001	.307
Error (G * L * T)	.101	28	.004			

Table 3: Anova for Same/Different Group Trials at Size 15

Source	*SS*	*df*	*MS*	*F*	*Sig*	*Partial Eta²*
Grouping	.019	1	.019	5.512	.026	.164
Error (Grouping)	.096	28	.003			
Location	.001	1	.001	.266	.610	.009
Error (Location)	.070	28	.002			
Trial Type	.118	1	.118	32.281	.000	.543
Error (Trial Type)	.099	28	.004			
Grouping * Location	.030	1	.030	13.170	.001	.320
Error (G * L)	.063	28	.002			
Grouping * Trial Type	.002	1	.002	.874	.358	.030
Error (G * T)	.089	28	.003			
Location * Trial Type	.001	1	.001	.275	.604	.010
Error (L * T)	.089	28	.003			
Grouping * Location * Trial Type	.023	1	.023	6.266	.018	.183
Error (G * L * T)	.101	28	.004			

Table 4: Anova for Same/Different Group Trials at Size 25

Source	*SS*	*df*	*MS*	*F*	*Sig*	*Partial Eta*^{2}
Grouping	.006	1	.006	2.197	.149	.073
Error (Grouping)	.080	28	.003			
Location	.001	1	.001	.307	.584	.011
Error (Location)	.102	28	.004			
Trial Type	.011	1	.011	2.334	.137	.077
Error (Trial Type)	.132	28	.005			
Grouping * Location	.002	1	.002	.763	.390	.027
Error (G * L)	.073	28	.003			
Grouping * Trial Type	.025	1	.025	5.538	.026	.165
Error (G * T)	.125	28	.004			
Location * Trial Type	.000	1	.000	.000	.987	.000
Error (L * T)	.096	28	.003			
Grouping * Location * Trial Type	.000	1	.000	.000	.984	.000
Error (G * L * T)	.099	28	.004			

Table 5: Anova for Singleton Singleton and Sister Singleton Group Trials

Source	*SS*	*df*	*MS*	*F*	*Sig*	*Partial Eta Squared*
Size	63.232	2	31.616	741.052	.000*	.964
Error (Size)	2.389	56	.043			
Grouping	.062	1	.062	7.989	.009	.222
Error (Grouping)	.218	28	.008			
Location	.003	1	.003	1.016	.322	.035
Error (Location)	.096	28	.003			
SG/GG Trial Type	.079	1	.079	17.261	.000	.381
Error (SG/GG)	.128	28	.005			
Size * Grouping	.014	2	.007	1.286	.284	.044
Error (S * G)	.308	56	.006			
Size * Location	.004	2	.002	.333	.718	.012
Error (S * L)	.314	56	.006			
Grouping * Location	.000	1	.000	.019	.893	.001
Error (G * L)	.174	28	.006			
Size * Grouping * Location	.002	2	.001	.197	.822	.007
Error (S * G * L)	.264	56	.005			
Size * SS/SG Trial Type	.011	2	.005	1.081	.335*	.037

Error (S * SS/SG)	.285	56	.005			
Grouping * SS/SG Trial Type	.004	1	.004	1.002	.321	.035
Error (G * SS/SG)	.110	28	.004			
Size * Grouping * SS/SG Trial Type	.016	2	.008	1.251	.294	.043
Error (S * G * SS/SG)	.348	56	.006			
Location * SS/SG Trial Type	.000	1	.000	.001	.979	.000
Error (L * SS/SG)	.224	28	.008			
Size * Location * SS/SG Trial Type	.006	2	.003	.541	.585	.019
Error (S * L * SS/SG)	.305	56	.005			
Grouping * Location * SS/SG Trial Type	.000	1	.000	.002	.967	.000
Error (G * L * SS/SG)	.104	28	.004			
Size * Grouping * Location * SS/SG Trial Type	.010	2	.005	1.150	.324	.039
Error (S * G * L * SS/SG)	.249	56	.004			

*corrected using the Huynh-Feldt correction due to the violation of the sphericity assumption

Table 6: ANOVA table of same group trial type

Source	SS	df	MS	F	Sig	Partial Eta^2
Size	34.942	2	17.471	566.304	.000*	.953
Error (Size)	1.728	56	.031			
Grouping	.203	1	.023	11.392	.002	.289
Error (Grouping)	.056	28	.002			
Location	.103	1	.103	29.584	.000	.514
Error (Location)	.097	28	.003			
Size*Grouping	.110	2	.055	12.793	.000	.314
Error (S*G)	.241	56	.004			
Size * Location	.142	2	.071	20.023	.000	.417
Error (S*L)	.199	56	.004			
Grouping * Location	.079	1	.079	26.596	.000	.487
Error (G * L)	.083	28	.003			
Size * Grouping * Location	.059	2	.029	7.879	.001	.220
Error (S*G*L)	.208	56	.004			

*adjusted using the Huynh-Feldt correction due to the violation of the sphericity assumption