



**Digital Commons@**

Loyola Marymount University  
LMU Loyola Law School

---

Honors Thesis

Honors Program

---

5-5-2023

## Comparing Hierarchical Data Structures and Hierarchical Data Analysis

Halley Jeanne Dante  
hdante@lion.lmu.edu

Robert Rovetti  
Loyola Marymount University, robert.rovetti@lmu.edu

Follow this and additional works at: <https://digitalcommons.lmu.edu/honors-thesis>



Part of the [Multivariate Analysis Commons](#), [Other Social and Behavioral Sciences Commons](#), and the [Statistical Models Commons](#)

---

### Recommended Citation

Dante, Halley Jeanne and Rovetti, Robert, "Comparing Hierarchical Data Structures and Hierarchical Data Analysis" (2023). *Honors Thesis*. 458.  
<https://digitalcommons.lmu.edu/honors-thesis/458>

This Honors Thesis is brought to you for free and open access by the Honors Program at Digital Commons @ Loyola Marymount University and Loyola Law School. It has been accepted for inclusion in Honors Thesis by an authorized administrator of Digital Commons@Loyola Marymount University and Loyola Law School. For more information, please contact [digitalcommons@lmu.edu](mailto:digitalcommons@lmu.edu).



# Comparing Hierarchical Data Structures and Hierarchical Data Analysis

A thesis submitted in partial satisfaction  
of the requirements of the University Honors Program  
of Loyola Marymount University  
in partial fulfillment of the requirements  
for Graduation with the Bachelor of Science Degree

by

**Halley Jeanne L. Dante and Robert Rovetti**

**May 2023**

Abstract

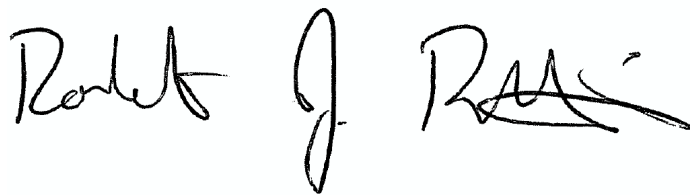
Real world data is inherently noisy and data analysis can be especially complex when noise is compounded in hierarchical and multilevel data structures. Since such data structures can be described using multiple approaches, the way data is collapsed and grouped within these structures can influence its resulting interpretation and analyses. To avoid discrepancies in data collapsing and grouping, multiple statistical approaches have been developed specifically to analyze multilevel data structures. Examples of multilevel statistical models are the two-factor ANOVA and the general linear model with repeated-measures (GLM-RR) which is typically used in the context of looking at change over time. Unlike simple summary-statistics such as t-tests, multilevel models allow for precision in the effect of each level on the observed data. In this study, analyses will be done using both simple statistical models and multilevel models with a dataset from a behavioral decision-making assay that aims to see whether phototactic preference changes over 24 hours in larval zebrafish. The simple and multilevel analyses will be compared through the descriptive analyses and hypothesis testing. The descriptive analyses will provide insight into the practicality of collapsing levels of data in hierarchical data structures and the hypothesis testing will provide comparative insight into the use of both simple and multilevel statistical models.

*Keywords:* Crossed Data Structures, Multilevel Modelling, Hierarchical Data

**Thesis written by**

Halley Jeanne L. Dante

Approved by



05 May 2023

---

Dr. Robert Rovetti, Thesis Director

Date



05 May 2023

---

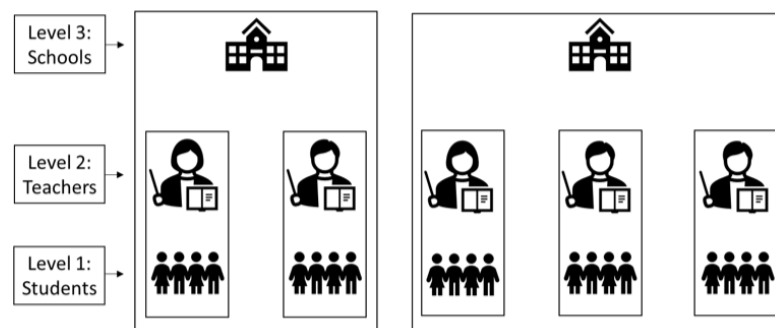
Dr. Lily Khadjavi, Mathematics Department Chair

Date

## Introduction

### *Hierarchical Data Structures*

Hierarchical data structures are a means of storing data in multiple levels, where each level pertains to a category associated with one aspect of an experimental design. Take, for instance, a study that aims to compare the pre-calculus test scores of two different schools as seen in *Figure 1*. In this example, the test scores are stored at the lowest level with the students, and each student is associated with specific teachers who are then associated with specific schools. The storing of data in lower-level categories which are nested within higher-level categories is the defining characteristic of hierarchical data structures.



*Figure 1: This figure represents an example of a hierarchical data structure with three levels: schools, teachers, and students. Figure taken from <https://www.statisticssolutions.com/what-is-hierarchical-linear-modeling/>*

To simplify the relationships that occur in hierarchical data structures, it is possible to mathematically represent such structures through nodes and links. *Figure 2A* is one example of this type of representation using the example of schools, teachers, and students. In this representation, the test scores are stored in students  $a$  through  $j$ , which each correspond to one unique teacher labeled 1 through 5, which each correspond to one of the two schools labeled A or B. In this simple example, each node on each level has a one-to-one relationship with nodes in other levels. In other words, each student is associated with a singular teacher, and each teacher is associated with a singular school.

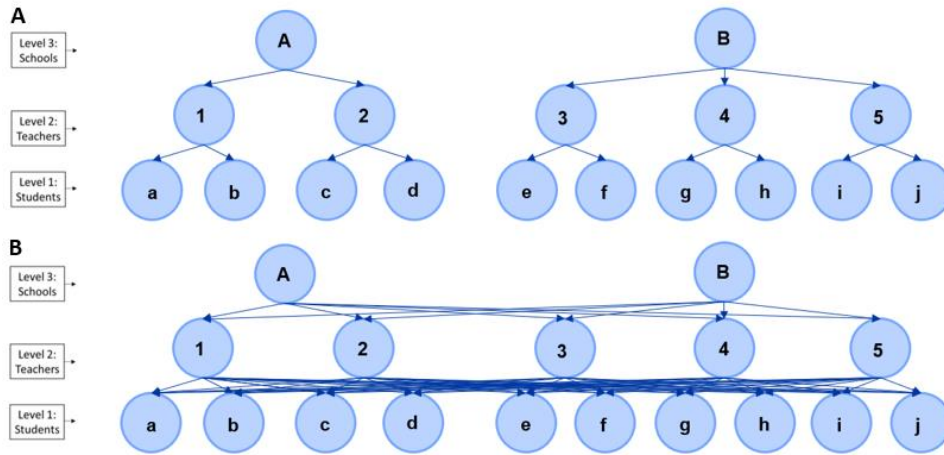


Figure 2: This figure represents the mathematical representation of the hierarchical data structure of schools, teachers, and students. Figure A) represents the simplest hierarchical data structure where every node has a one-to-one relationship and figure B) represents the completely crossed hierarchical data structure where every node is related to all other nodes on the level above and/or below it.

In the real world, more complex relationships can occur where not every relationship is a one-to-one relationship. For instance, a student can be enrolled under more than one teacher, and a teacher can be affiliated with more than one school. This crossing can continue until completion in which case all students are enrolled under all teachers, who are affiliated with both schools. This completely crossed hierarchical data structure is depicted in *Figure 2B* where every student has arrows that connect them to every teacher, and every teacher has arrows that connect them to both schools.

### *Analysis of Hierarchical Data Structures*

Since the nested nature of hierarchical data structures adds a layer of complexity to data analysis due to potential interactions between the different levels of data, the analysis of such structures needs to be done carefully. One way to do so is by applying multilevel modeling such as the two-way ANOVA and General Linear Models with repeated regression (GLM-RR). The two-way ANOVA tests for the effects of multiple independent variables on dependent variables, including the way the independent variables interact with one another (Hayes, 2022). For

hierarchical data structures, the independent variables would be the levels of categories and the dependent variables would be the recorded data.

Similar to the two-way ANOVA, a GLM-RR analyzes a dependent variable that is measured more than once for each subject. In this case, each measurement differs in its independent variables which are called within-subject factors. In hierarchical data structures, these factors are the levels of categories in which data is stored. The primary difference is that GLM-RR transforms the dependent variables according to their within-subject factors prior to analysis, whereas a two-way ANOVA analyzes the original dependent variables according to their within-subject variables (Wolfinger and Chang, 1996).

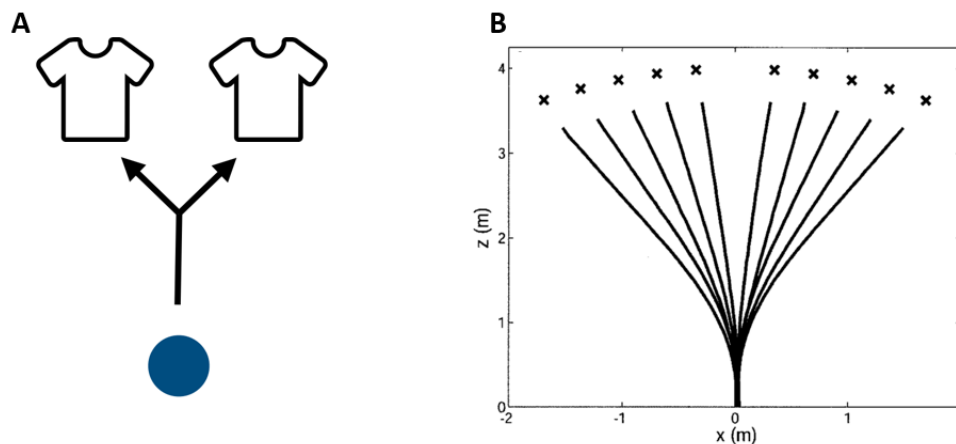
Despite the availability of analysis methods for multilevel data, one study found that traditional summary-statistics suffice for nested data. In this study, they find that multilevel models may suffer from frequent singular fit errors compared to simpler analyses (McNabb and Murayama, 2021). Thus, they suggest that running summary-statistics analyses like t-tests with clustered and collapsed data may suffice for hierarchical data structures. Doing so would require collapsing data into one category and analyzing it accordingly, which also requires less computing power due to its relative simplicity compared to two-way ANOVA and GLM-RR.

Due to the variety of analysis methods that are available for hierarchical data structures, one of the objectives of this study is to compare the descriptive and statistical analyses of these methods. In addition, since summary-statics require the clustering and collapsing of data into one level, this study aims to compare the descriptive and statistical results of different collapsed data structures.

### *Spatial Decision-Making*

In this study, the hierarchal data structure used was collected from a behavioral experiment that explored spatial decision-making. Spatial decision-making involves the intake of stimulus choices, the internal processing of a stimulus, the process of deciding, and a resulting output that involves movement (Eldrandaly, 2011). In the real world, one example of this is when one goes clothes shopping and is choosing to decide which stores to shop in. In this case, the store displays are visual stimuli choices, the internal processing involve an interplay with the stimuli and your personal preference, the process of deciding occurs when you choose a shop to enter, and the output involves your actual movement towards your preferred shop.

Due to the nature of spatial decision-making, decisions can be measured via spatial information. *Figure 3* depicts one of the representations of spatial decision-making which is represented via the potential decision paths one can take. *Figure 3A* depicts the earlier example of going clothes shopping which is measured by observing the two decision paths that can result from choosing either of the two visual choices. The more choices one is presented with, the more decision paths one can take which is represented in *Figure 3B* (Fajen and Warren, 2003).

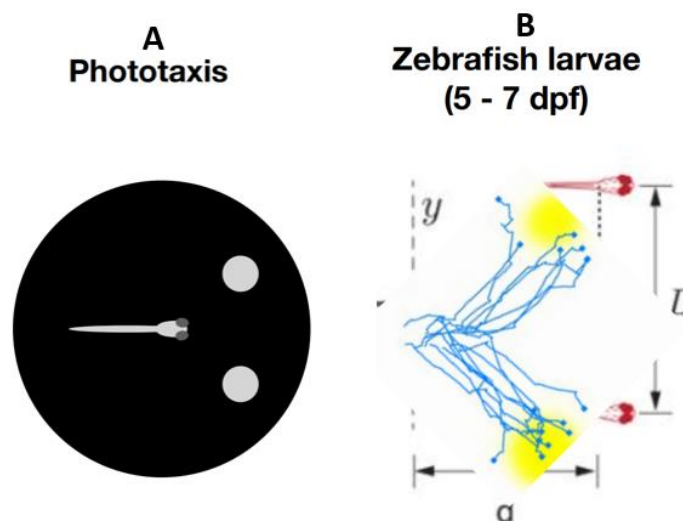


*Figure 3: This figure depicts the mathematical representation of spatial decision-making. Figure A) represents spatial decision-making with two visual choices and two resulting decision paths while figure B) represents spatial decision-making with ten visual choices and ten resulting decision paths.*



### *Phototaxis in Larval Zebrafish*

The spatial decision-making data used in this study was collected by utilizing phototactic behaviors in larval zebrafish. By definition, phototaxis is a behavior that involves movement in response to light stimuli where positive phototaxis is movement towards light and negative phototaxis is movement away from light. In larval zebrafish, it is known that positive phototactic behavior is displayed (Chen and Engert, 2014). One study utilized this behavior in their experimental design to study spatial decision-making where they presented two light stimuli to a fish on a black background as depicted in *Figure 4A*. In response to the stimuli, fish swam towards the light and underwent one of two decision paths as depicted in *Figure 4B* (Burgess *et al.*, 2010). The movement that the larval zebrafish underwent mirrors the mathematical representations of spatial decision-making in *Figure 3*. This suggests that the use of light stimuli as visual stimuli for spatial decision-making paradigms in larval zebrafish is justifiable due to their positive phototactic behaviors.



*Figure 4: This figure depicts the use of phototaxis in spatial decision-making behavioral paradigms. Figure A) depicts the stimuli choices in such experimental paradigm and figure B) depicts the resulting movements of larval zebrafish. The movement of larval zebrafish in figure B mirrors the mathematical representation of spatial decision-making, which supports the use of phototactic stimuli to study spatial decision-making*

### *Timescales and Flexible Phototaxis*

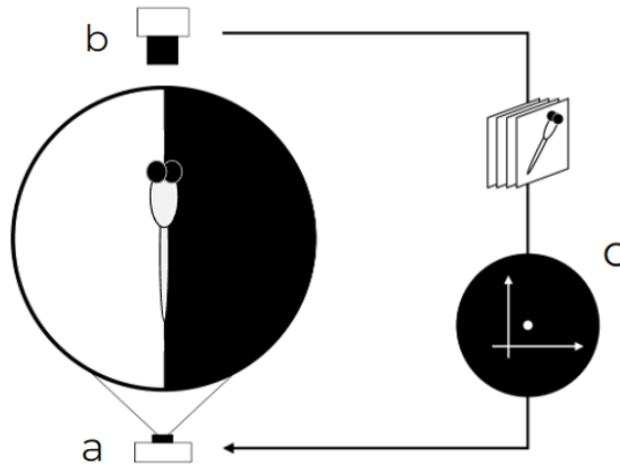
Though larval zebrafish have been shown to display positive phototaxis, recent studies have found that these behaviors undergo fluctuations over different timescales. These timescales have been found to occur over short time scales such as when zebrafish move in and out of shaded areas, and long timescales such as when the sun rises and sets (Chen *et al.*, 2021).

Depending on the history of a zebrafish's light exposure, the strength of phototactic behaviors can change. Due to this finding, the primary experimental question of this study is to determine whether phototactic behaviors change over 24 hours and if so, whether that change is consistent across different light stimuli pairs.

## **Methods**

### *Experimental Setup and Data Collection*

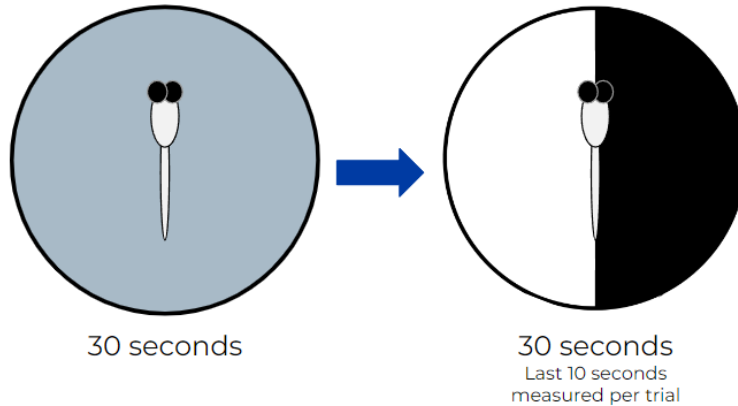
In this study, data was collected through a closed-loop stimulus setup which is depicted in *Figure 5*. As shown, the setup consists of four parts: a) a projector that displays a split view stimulus that appears half-light and half-dark onto the bottom of a petri dish, b) a camera that tracks the orientation and location of the larval zebrafish, and c) a computer that translates the positional data from the camera back into the projector. By tracking and translating the positional data in real-time, the setup is able to adjust itself based on the zebrafish's location in the dish and adjust the stimulus accordingly. Doing so allows the setup to precisely control what is being projected onto each eye of the fish regardless of where it is on the dish and which direction it is facing. This way, spatial decision-making can be measured by counting how many light-ward turns a fish takes and how many dark-ward turns a fish takes.



*Figure 5: This figure depicts the experimental setup of the closed-loop behavioral assay where a. shows the projector which displays the stimuli to the bottom of the petri dish; b. shows the petri dish that contains a larval zebrafish; c. shows the camera that tracks and collects positional data of larval zebrafish which computes the positional data and feeds it back to projector. Through this loop, the placement of the stimulus responds to the movement of the fish so that the light stimulus is fixed to one eye throughout the trial period, and the dark stimulus is fixed to the other eye.*

In general, a trial consists of 1-minute trials that consists of a 30 second adaptation period and a 30 second experimental period as depicted in *Figure 6*. Data collection occurs during the experimental period where the stimulus appears as a split-view which consists of two different light intensities ranging from pure black to pure white. Due to the closed-loop nature of the setup, each of the light intensities is fixed onto either eye of the fish. This makes it so that the left eye sees only one light intensity throughout the duration of the experiment and the right eye sees the other light intensity throughout the duration of the experiment regardless of the orientation and direction of the fish. The light intensity shown to each eye is randomized every trial so that both eyes are exposed to both bright and dark intensities. This randomization controls for any turning bias that the fish might have. During the adaptation period, the stimulus appears as a homogenous background which consists of the average light intensity of the following

experimental stimuli. This adaptation period is designed to control for any short-term bias that may be carried over from the previous experimental stimuli.



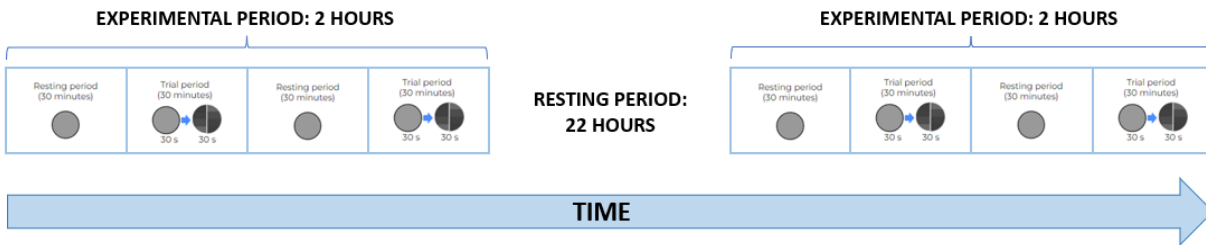
*Figure 6: This figure shows the experimental design of the different stimuli shown during the adaptation period (left) and experimental period (right). The adaptation period consists of a 30-second stimulus and the experimental period consists of a 30-second split-view stimulus.*

Regarding data collection, the raw measurement in this experiment is the count of turns a fish takes towards the lighter intensity in relation to the total turns taken during the trial. For the analysis of this study, these raw turns were converted into a percentage which represents lightward turns. Since the shift from the adaptation stimuli to the experimental stimuli may result in noisy behaviors, the only turns that were included in this experiment were those that took place during the last 10 seconds of the experimental period when turning stabilized.

### *Experimental Design*

The primary experimental design of this study consisted of two controls: time and stimulus intensity. Regarding time, data collection occurred over 1 hour experiment blocks that consisted of a 30-minute resting period and a 30-minute trial period. During the 30-minute resting period, the stimulus displayed a homogeneous grey background which was intended to control for any long-term variation in light exposure that the fish may have had prior to the experiment. During the 30-minute trial period, the stimulus displayed 30 1-minute trials, where

each trial consisted of the 30-second adaptation period and a 30-second trial period described previously. This experiment block was repeated twice, followed by a 22-hour resting period where fish were left on petri-dishes and exposed to light during daytime and dark during nighttime. After 22 hours of resting, the fish were then exposed to another 2 experiment blocks. This temporal design is summarized in *Figure 7*.

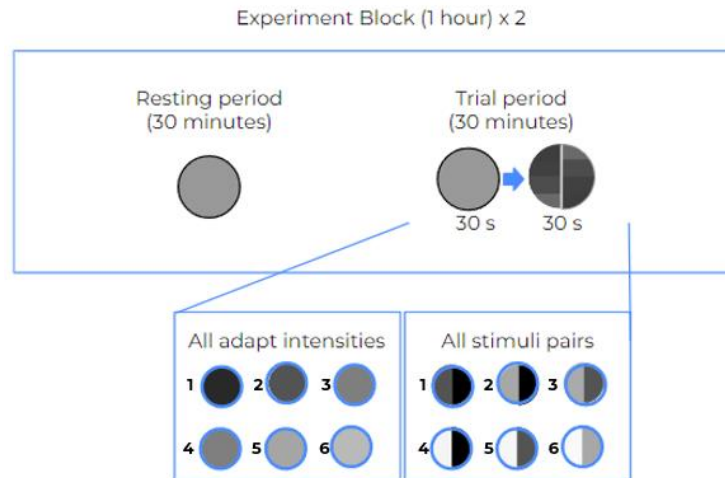


*Figure 7: This figure shows the temporal components of the experimental designs. The components include 1) the 26-hour timescale which involve 2 hours of data collection, 22 hours of rest, and 2 hours of repeated data collection, 2) the 2-hour experimental periods which are composed of two rounds of 30-minute resting periods and 30-minute trial periods, and 3) the trial periods which consist of 30-second adaptation periods and 30-second experimental periods.*

For the entire duration of data collection, 32 zebrafish were used and their fish IDs were tracked to ensure that the experimental results could be traced to an individual level and that data could be paired. This, alongside the temporal design of the study, allowed for the experiment to see whether phototactic preference changed over the span of 24 hours.

Regarding stimulus intensity, there were six light intensity pairs that the fish were exposed to during the trial period as depicted in *Figure 8*. These pairs were different combinations of four light intensities which were complete white, light grey, dark grey, and complete black. The intensities were specified on a digital greyscale where black was 0, dark grey was 0.3, light grey was 0.6, and white was 1. During the adaptation period prior to the trials, the light intensity shown was the average of the numerical value associated with the trial light intensity pairs. Throughout the 30-minute trial period, the stimuli pairs were randomized for all fishes. This light intensity design allowed for the experiment to see whether phototactic

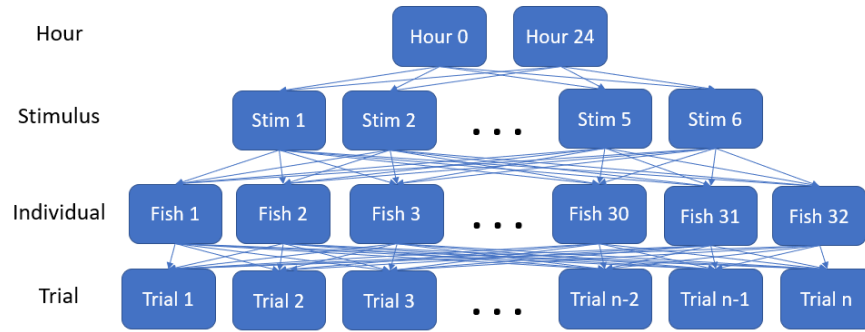
preference was dependent on the type of stimuli the fish were presented with. Pairing the stimuli this way also allows for the experiment to detect whether specific characteristics of the stimuli such as the difference in the stimuli intensities, the darker stimuli intensity, or the lighter stimuli intensity are more likely to influence phototactic preference and its persistence over time.



*Figure 8: This figure shows the experimental design of the different stimuli pairs that were shown. The intensity values are as follows: 1) adaption intensity of 0.183, light intensity of 0.33, dark intensity of 0, 2) adaption intensity of 0.33, light intensity of 0.66, dark intensity of 0, 3) adaption intensity of 0.495, light intensity of 0.66, dark intensity of 0.33, 4) adaption intensity of 0.500, light intensity of 1, dark intensity of 0, 5) adaption intensity of 0.66, light intensity of 1, dark intensity of 0.33, and 6) adaption intensity of 0.83, light intensity of 1, dark intensity of 0.66*

### Data Structure

The primary question that this paper aims to answer regarding the dataset is whether phototactic preference changes over 24 hours. However, since there were other variables that were modified and controlled for in the experiment including the six stimulus pairs, each individual fish ID, and the trial number as recorded by the computer, the structure of the dataset is hierarchical and crossed as summarized in *Figure 9*. Essentially, the recorded data of the raw count of light-ward and dark-ward turns a fish takes in the last ten seconds of a trial is recorded for each trial. Each trial is associated with a specific fish, a specific stimulus, and a specific hour.



*Figure 9: This figure reveals the hierarchical data structure of the phototactic, spatial decision-making data from larval zebrafish. At the lowest level, the light-ward turns are measured and stored at the trial level, which are then stored at the individual fish level, which are then stored at the stimulus pair level, which are then stored at the hour level.*

Due to the hierarchical nature of the data structure, there are multiple approaches to data analysis that can be done in order to answer whether phototactic preference changes over 24 hours including simple linear models alongside multilevel models. Therefore, the second question this paper aims to answer is to compare the statistical results of simple-linear models and multilevel models. This will first be done with descriptive analysis and the comparison of means, followed by statistical analysis and the comparison of statistical results. After a comparison is done, a closer assessment of the nature of the data will be conducted in order to potentially explain the differences or similarities that were observed.

Since each type of model requires different data structures, the data used for analysis collapsed in different ways. *Table 1* shows the three ways of collapsing the hierarchical data structure into one level. To explain how the collapsing is done, all the raw data that were collected under a specific variable were grouped together in counts of left turns and counts of right turns. For example, the turn counts from all fishes and all trials that were done using stimulus 1 on hour 0 were grouped together and the turn counts from all fishes and all trials that were done using stimulus 1 on hour 24 were grouped together. This was done for all stimuli on both days and would be considered the collapsing of the data structure into the stimulus level.

Then, the counts were converted into a percentage which would represent how often fishes turned towards the lighter intensity during the experiment. In this paper, this measure will be referred to as percent light-ward turn (LT%). This collapsing of raw counts and conversion into a percentage was done for all the variables: stimulus, individual fish, and trial. Throughout this paper, this collapsing will be referred to as one-level, one-variable data structures.

*Table 1: This table shows the three different ways that data was collapsed in this study. At the stimulus level, all turn counts regardless of fish and trial number were collapsed into each stimulus giving a total of 6 datapoints per hour, where each datapoint consisted of an average of 10065 total turns. At the individual level, all turn counts regardless of stimulus and trial number were collapsed into each individual fish giving a total of 32 datapoints per hour, where each datapoint consisted of an average of 1887 total turns. At the trial level, all turn counts regardless of individual fish and stimulus were collapsed into each trial number giving a total of 6 datapoints per hour, where each datapoint consisted of an average of 17 total turns.*

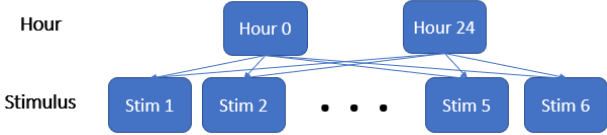
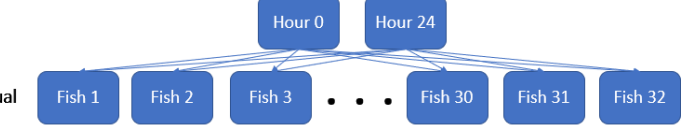
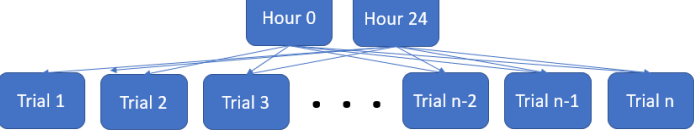
Level	Data Structure	Sample Size (n) and Average Denominator (d)
Stimulus		n=6  d=10065
Individual		n=32  d=1887
Trial		n=3600  d=17

Table 2 shows the two ways of collapsing the hierarchical data structure that was used for multi-level analysis. Similar to the one-level data structures, the turn counts were collapsed into the groups described by the lowest level of the data structure, followed by a conversion into a percentage. Both data structures for multi-level analysis took into account data from the different



stimuli and data from each individual fish. The only difference between the one-level data structure and the two-level data structure is that the one-level data structure contained the turn counts for each combination of stimulus and individual fish. Since there are 6 stimuli and 32 individual fish, this meant that there was a total of 192 measurements under this data structure. Throughout this paper, this collapsing will be referred to as one-level, two-variable data structures. In contrast, the two-level data structure contained the turn counts for each individual fish, meaning that there was a total of 32 measurements. However, each measurement was associated with two other categorical variables: hour and stimulus. Throughout this paper, this collapsing will be referred to as two-level data structures.

Table 2: This table shows the two ways that hierarchical data was analyzed in this study. The one-level data structure consisted of 192 datapoints per hour which averaged 315 total turn counts. The datapoints in this structure are stored at the level which represents unique pair of individual fish and stimulus. The two-level data structure consisted of 32 datapoints per hour which averaged 1887 total turn counts. The datapoints in this structure are stored at the individual fish level which are nested in unique stimuli.

Level	Data Structure	Sample Size (n) and Average Denominator (d)
One-level: Individual/ Stimulus	<p>The diagram shows a hierarchical structure. At the top level, there are two boxes labeled 'Hour 0' and 'Hour 24'. Below these, there are two levels of boxes representing 'Stimulus/Individual'. The first level has boxes for 'Stim 1 / Fish 1', 'Stim 1 / Fish 2', 'Stim 1 / Fish 3', followed by three dots, then 'Stim 6 / Fish 30', 'Stim 6 / Fish 31', and 'Stim 6 / Fish 32'. Arrows point from each 'Hour' box to each 'Stimulus/Individual' box, indicating that each hour contains data for all stimulus-fish combinations.</p>	n=192  d=315
Two-level	<p>The diagram shows a hierarchical structure. At the top level, there are two boxes labeled 'Hour 0' and 'Hour 24'. Below these, there is a level of boxes representing 'Stimulus' (Stim 1, Stim 2, followed by three dots, Stim 5, Stim 6). Arrows point from each 'Hour' box to each 'Stimulus' box. Below the 'Stimulus' level, there is a level of boxes representing 'Individual' (Fish 1, Fish 2, Fish 3, followed by three dots, Fish 30, Fish 31, Fish 32). Arrows point from each 'Stimulus' box to each 'Individual' box, indicating that each stimulus is measured for all individual fish.</p>	n=32  d=1887

### *Data Analysis*

The comparison of the different data structures occurred in two ways: through descriptive analysis and statistical analysis. The descriptive analysis will simply be the comparison of the means of hour 0 and hour 24 for each data structure. For the comparison of means of the two-level data structures, this will be done through two different approaches: a two-factor ANOVA alongside a general linear model with repeated measures (GLM-RR). This will then be followed by statistical analysis which will be done with simple paired t-tests for one-level data structures and a two-factor ANOVA and GLM-RR for the two-level data structure. In this study, the t-tests were run in Excel and the multilevel analyses were run in SPSS.

## **Results and Discussion**

### *Descriptive Analysis*

*Figure 10* summarizes the mean LT% for hour 0 and hour 24 across different data structures and across different multilevel descriptive summaries. As depicted, the mean LT% for the one-variable data structures differed from one another where the trial data structure, the individual data structure, and the stimulus data structure had hour 0 means of 51.6%, 51.1% and 51.4% respectively, and had hour 24 means of 53.8%, 54.3%, and 53.1% respectively. This leads to a range of 0.5% for the hour 0 means and a range of 1.2% for the hour 24 means. In comparison, the mean LT% for the two-variable data structures were more similar to one another where the one-level individual/stimulus data structure, the two-factor ANOVA summary, and the GLM-RR summary had hour 0 means of 51.3%, 51.3%, and 51.4% respectively, and had hour 24 means of 54.4%, 54.4%, and 54.4% respectively. Compared to the one-variable means, the two-variable means had a range of 0.1% for hour 0 and 0.0% for hour 24.

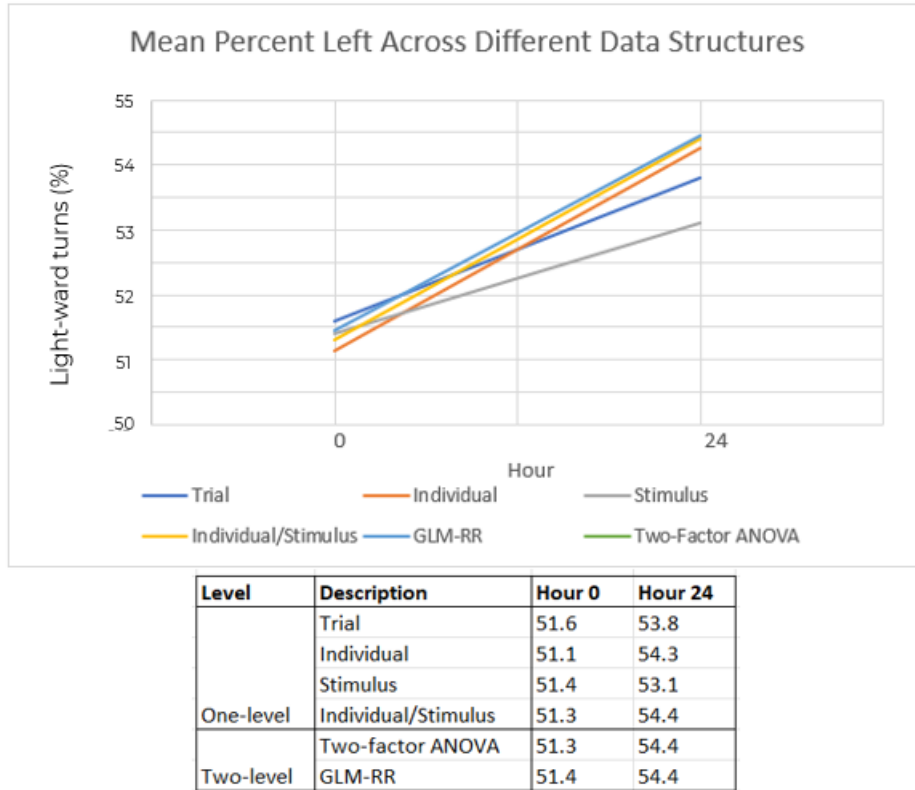


Figure 10: This figure reveals the descriptive results of the different data structures. The graph above shows the plots of the means using the four different collapsed data structures alongside the two different multilevel models for both hour 0 and hour 24. The table below shows the raw results of the graph for both hour 0 and hour 24.

Comparing the means of the different data structures, all show a general upward trend which suggests that there is an overall increase in positive phototactic behavior over 24 hours in larval zebrafish. Despite this similarity, there are two data structures which have different slopes of %LT over 24 hours which are the trial-level data structure and the stimulus-level data structure. The means of these data structures are depicted in the blue and grey lines respectively in the graph in *Figure 10*.

To assess why the trial-level and the stimulus-level show a slightly different slope from the other data structures, a closer look at the nature of the different distributions is required.

*Figure 11* summarizes the distribution and the sample size of %LTs for each one-level data structure. As depicted, the sample sizes of the trial, individual, stimulus, and individual/stimulus

variables were 3600, 32, 6, and 192 fishes respectively. Comparing the sample sizes, the small sample size of the stimulus variable may explain why the stimulus-level means differed from the two-level data structures since smaller sample sizes are more prone to sampling bias. Comparing the distributions, the distribution of the trial variable may explain why the trial-level means differed from the two-level data structures since the high variability in the data can introduce more noise.

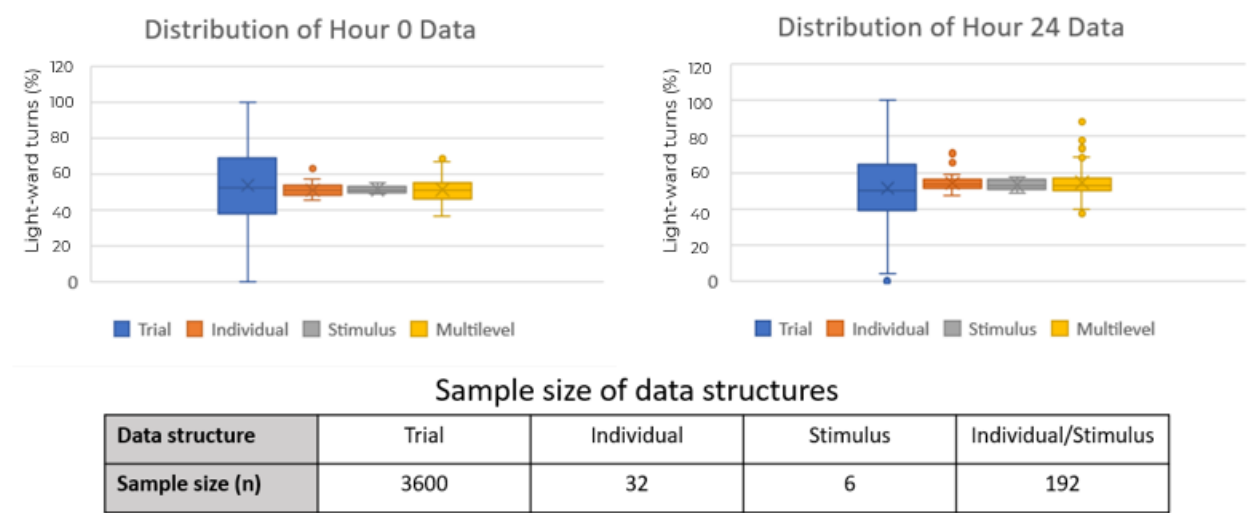
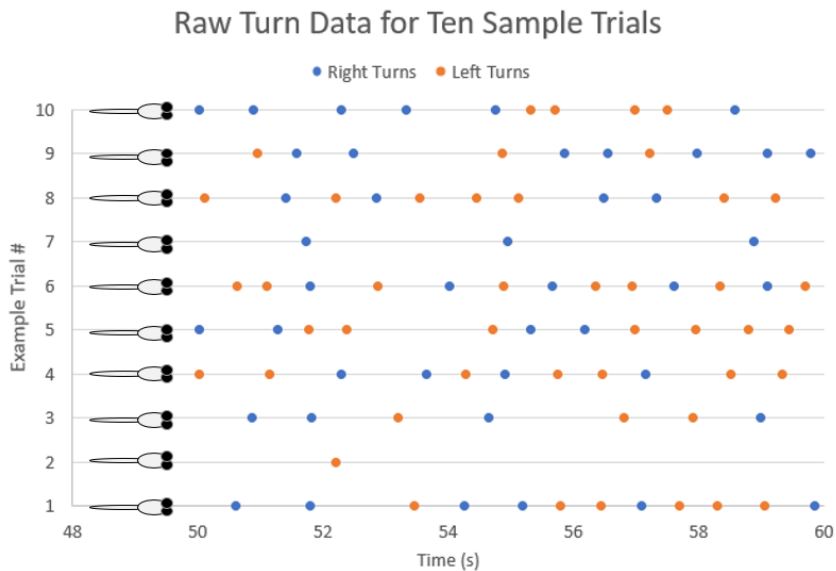


Figure 11: This figure shows the distributions and the sample sizes of the %LTs for all four collapsed one-level data structures. The distribution graphs reveal a larger than average range of %LTs for the trial-level and the sample size table reveals a smaller than average sample size for the stimulus-level.

The high variability of the trial-level data may be attributed to the fact that the trial-level means had lesser turning counts compared to other one-level variables. This means that it may be prone to sampling bias since lesser turning counts allow for extreme %LTs to become more likely. For instance, if a fish happened to turn left once in a given trial and did not turn right at all, then the raw left turn count would be 1 while the %LT would be 100%. However, if a fish happened to turn left once in a given trial and turned right once, the raw left turn count would still be 1, but the LT% would drop to 50%. Since the weight of every trial's %LT is treated

equally in the calculation of the means, the weight of the turns themselves would differ and the trials with smaller turn counts are more likely to skew the mean %LT. This is seen in *Figure 12* which gives sample trial data for a random selection of fish. The effect of small sample size can be seen best with fish 2 which has only one turn for this trial, making it's %LT 100%. Due to the large distribution range of the trial-level data structure and the small sample size of the stimulus-level data structure, the deviating means may be due to these abnormalities.



*Figure 12: This figure shows a sample of the raw turn data from 10 trials over the span of the last 10 seconds where data was collected. In the graph, orange depicts left turns and blue depicts right turns. As observed, the number of turns taken per trial can vary.*

*One-Level Statistical Analysis – Comparing Collapsed One-Level Data Structures*

*Table 3* summarizes the paired t-test results of the one-level data structures with the following hypothesis:

$$H_0: \mu_{\text{hour 0 PL\%}} = \mu_{\text{hour 24 PL\%}}$$

$$H_A: \mu_{\text{hour 0 PL\%}} \neq \mu_{\text{hour 24 PL\%}}$$

Similar to the descriptive analysis result, the p-values of the one-variable means differed from one another. Specifically, the p-value of the trial-level, individual-level, and stimulus-level were 0.000, 0.013, and 0.057 respectively. The paired t-test that differed the most from other paired t-tests was the one

conducted with the stimulus-level. Like the descriptive analysis, this may be due to the stimulus variable’s small sample size.

*Table 3: This table shows the change in means alongside the paired t-test results of all the one-level data structures including the trial-level, individual-level, stimulus-level, and the individual/stimulus-level data structures. \*\* depicts significance at an  $\alpha=0.05$  level and \* depicts significance at an  $\alpha=0.01$  level*

Level	Description	$\Delta$ mean	df	Sig (2-tailed)
One-level	Trial	2.21%	3599	0.000**
	Individual	3.12%	31	0.013*
	Stimulus	6.90%	5	0.057
	Individual/Stimulus	3.08%	191	0.000**

As observed in the table, the p-values were overall low, with the individual-level and stimulus-level data structures being relatively larger. Given the sample size of the individual-level variable (n=32) which is smaller than the trial-level (n=3600) and the individual/stimulus-level (n=191), it is likely that the larger p-value may be due to a relatively smaller sample size. This can also be seen in the stimulus-level which has a small sample size (n=5). This suggests that when running statistical tests on collapsed data, paying attention to the sample size of the level that the data is collapsed to is important to consider as it may have an effect on the resulting p-value.

*Two-Level Statistical Analysis – Comparing ANOVA and GLM-RR*

The results for the two-factor ANOVA and the GLM-RR are summarized in the table of *Figure 13* which shows similar relatively small p-values for the effects of all levels. In other words, it appears to show that there is a statistically significant effect of the day on the %LT and that there is a statistically significant effect of the stimulus on the %LT. However, there appears to be a larger p-value for the interaction term of stimulus and day, suggesting that the effect of the day on %LT may not be stimulus specific. This is shown in the bar graph and the plot of the two-way ANOVA and GLM-RR results in *Figure 13*. Especially in the plot of %LT over 24

hours across the different stimuli, there seems to be an upward trend regardless of the stimulus. This supports the larger p-value of the interaction term.

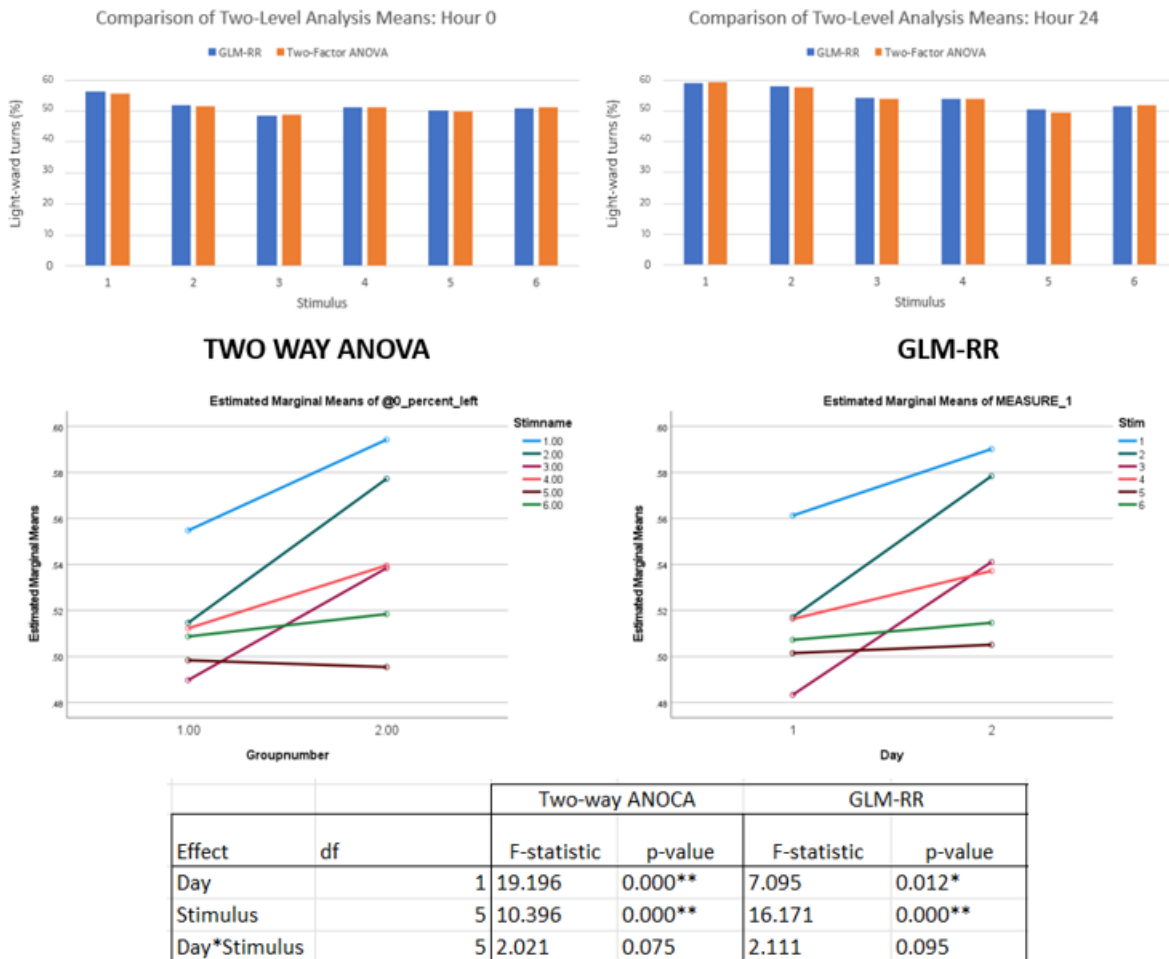


Figure 13: This figure shows the results of the two-way ANOVA and the GLM-RR analyses of the two-level hierarchical data structure. The top graph depicts the bar graphs of the different mean %LT of each stimuli pair for both tests on hour 0 (left) and hour 24 (right). The middle graph depicts the plots of the different mean %LT of each stimuli pair for hour 0 and hour 24 for both the two-way ANOVA (left) and the GLM-RR (right). The table at the bottom depicts the numerical results of both statistical tests including the F-statistic and its corresponding p-value. \*\* depicts significance at an  $\alpha=0.05$  level and \* depicts significance at an  $\alpha=0.01$  level

Comparing the statistical results of the two-way ANOVA and the GLM-RR, the p-values of both the day effect and the day\*stimulus effect appears to be relatively larger in GLM-RR. To assess why this might be, a closer look at the GLM-RR model is needed. Since GLM-RR

analyzes hierarchical data based on transformed data, it assumes sphericity where the variances of all combinations of related groups are equal. This assumption is measured by Mauchly’s test of sphericity, and any p-values lower than a 0.05 violates this assumption. As depicted in *Table 4*, the day effect has no p-value due to it having 0 degrees of freedom. However, the p-values of the stimulus effect and the day\*stimulus effect have p-values of 0.036 and 0.019, which violate the assumption. This violation may be the underlying reason why a larger p-value was observed for GLM-RR compared to the two-way ANOVA.

*Table 4: This table shows the Mauchly’s test of sphericity for the GLM-RR test run using the two-level hierarchical data structure*

**Mauchly's Test of Sphericity<sup>a</sup>**

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>b</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Day	1.000	.000	0	.	1.000	1.000	1.000
Stim	.424	24.954	14	.036	.737	.849	.200
Day * Stim	.394	27.083	14	.019	.756	.874	.200

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

**Conclusion**

In conclusion, there appears to be an overall increase in the positive phototactic behaviors of larval zebrafish regardless of how the hierarchical data structure is collapsed and which statistical test is used. Regardless, there were differences in the calculated change in mean %LT over 24 hours alongside differences in the calculated p-values associated with these differences. For mean %LT observed in the descriptive analysis, these differences can potentially be due to small sample sizes and large distributions as observed in the data structures that were collapsed into the stimulus-level and trial-level. In this study, the large range of data distribution may be due to the small total turns taken at the trial level which allows for the increased likelihood of extreme datapoints. In this case, both the large data distribution range and the small sample size are inherently sample size issues. This implies the importance of sample size of the datapoints



within a given level and the measurements taken within a given recorded datapoint when determining which level to collapse a hierarchical data structure into.

For the p-values observed in the statistical analysis, it appears that sample size also plays a role where smaller sample sizes lead to larger p-values in collapsed data structures. However, differences in observed p-values of multilevel statistical analysis may be due to other factors. In this study, a test-specific assumption for GLM-RR was violated, namely the Mauchly's test of sphericity. This violation may have led to the observation of a larger p-value compared to the two-way ANOVA. This reveals the importance of choosing the right multilevel test, and meeting test-specific assumptions as these may lead to differences in p-values.

Overall, this study shows that it is possible for simple summary statistics to reveal similar trends in collapsed data structures as multilevel models. This finding is useful in cases that primarily look to compare two means with the most efficient computing power. However, it is important to consider sample size of both the recorded data and the level at which data is collapsed into when running these simple summary statistics. Though efficient for computing power, this may not be a useful method of analysis if a study is interested in looking at the interplay of multiple effects on a dependent variable. In this case, a multilevel analysis is required, but special attention must be made to the assumptions of the multilevel model used. Any violations to the assumption may impact the p-values and the subsequent interpretation of these effects.

## References

- Burgess, H. A., Schoch, H., and Granato, M. (2010) Distinct retinal pathways drive spatial orientation behaviors in zebrafish navigation. *Curr Biol* 20(4):381-386.  
<https://doi.org/10.1016/j.cub.2010.01.022>
- Chen, X. and Engert, F. (2014). Navigational strategies underlying phototaxis in larval zebrafish. *Front. Syst. Neurosci.* 8(39). <https://doi.org/10.3389/fnsys.2014.00039>
- Chen, A. B., Deb, D., Bahl, A. and Engert, F. (2021). Algorithms underlying flexible phototaxis in larval zebrafish. *Journal of Experimental Biology.* 224(10).  
<https://doi.org/10.1242/jeb.238386>
- Eldrandaly, K. A. (2011). *Geographic Information Systems: GIS and Spatial Decision Making*. New York: Nova Science Publishers.
- Hayes, R. (2022). *Two-Way ANOVA: What it is, what it tells you, vs. One-way ANOVA*. Investopedia. Retrieved online at <https://www.investopedia.com/terms/t/two-way-anova.asp>
- McNabb, C. B. and Murayama, K. (2021). Unnecessary reliance on multilevel modelling to analyze nested data in neuroscience: When a traditional summary-statistics approach suffices. *Current Research in Neurobiology.* 2.  
<https://doi.org/10.1016/j.crneur.2021.100024>
- Wolfinger, R.D., & Chang, M. (1996). Comparing the SAS® GLM and MIXED Procedures for Repeated Measures.