Characterizing the Dynamics of Movement in *Drosophila Melanogaster* and Humans With Autistic Spectrum Disorder

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37364723">https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37364723</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Characterizing the dynamics of movement in *Drosophila melanogaster* and humans with Autistic Spectrum Disorder

A thesis submitted by

Ellen Li

To

the Department of Applied Mathematics

in partial fulfillment of the honors requirements

for the degree of

Bachelor of Arts

Harvard College

Cambridge, Massachusetts

April 3, 2020
Abstract

In this thesis, I seek to understand the dynamics and geometry behind two forms of movement by formulating a generative model that encapsulates key features of the movement. The first form of movement involves optogenetically stimulated *Drosophila melanogaster* walking in a two-dimensional virtual reality environment with a grid of obstacles. This virtual paradigm allows us to analyze the dynamics of the flies’ movement and identify two major modes of movement: exploring and traveling. We demonstrate that these modes of movement serve as the underpinnings of trends involving speed, angular velocity, and curvature that we observe in the data. The second form of movement involves the tracing capabilities of people with Autistic Spectrum Disorder (ASD), for whom research has revealed significant differences in motor movement as compared to the neurotypical population, but the mechanisms driving these differences are largely unknown. We examine the tracing movements of both groups and identify tracking delay as a possible mechanism for this disparity in the tracing task. For both of these forms of movement, we develop a generative model informed by our analysis of the data and demonstrate its validity by replicating identified trends in a simulated data set. This suggests that the key features we identify in both scenarios play a role in governing the dynamics of the two forms of movement.
Acknowledgements

First and foremost, I give my heartfelt gratitude to my thesis advisor, Professor L. Mahadevan. He has guided me in so many ways during both the research and thesis-writing process, and I know I can always depend on him to offer clarity, purpose, personal and academic inspiration, and a deluge of advice and wisdom that have guided me through my research.

I would also like to thank Professor Yuval Hart, who has provided much-needed guidance and acumen, particularly during the early parts of my research journey. He taught me the basics and essentials of research, from approaching a problem for the first time to collecting and presenting one’s thoughts and results towards the end of the project. Also, I would like to extend my appreciation to Professor Vivek Jayaraman and Hannah Haberkern at HHMI Janelia, both of whom introduced me to their fascinating research and provided wonderful insight during this process. This thesis could not have been possible without the guidance of these research mentors.

I also offer my sincere thanks to the Applied Math department here at Harvard for wonderful conversations, support, and advice throughout the course of my college career. In particular, Professors Yiling Chen, Sarah Iams, and Margo Levine have been constant sources of advice that have much improved my college experience.

On a more personal level, I would like to thank my blockmates and friends for their unending companionship and support during my college years. They have supported me through the best and worst of times with ceaseless laughter and constant friendship, offering a great deal of wisdom in all walks of life. I cannot thank them enough for shaping my college experience into the best four years of my life.

Finally, and most importantly, I would like to thank my parents for twenty-one years of guidance and unconditional love. From a young age, they have instilled in me a love of learning and have always encouraged me to strive for excellence, shaping me into the person I am today.
Contents

1 Introduction 13

2 Walking Behavior of Drosophila melanogaster 15
   2.1 Related Work 16
   2.2 The Data 17
   2.3 Initial Analysis 21
   2.4 Identifying Two Modes of Movement 24
   2.5 A Generative Model 29
   2.6 Discussion 33

3 The Motion of Tracing in People with ASD 37
   3.1 Related Work 38
   3.2 Tracing Data 41
   3.3 Curvature Calculation 44
   3.4 Curvature Alignment Theory 46
   3.5 Curvature Alignment of the Data 49
   3.6 Curvature Analysis 51
   3.7 Karcher Mean Analysis 55
   3.8 Speed Analysis 57
   3.9 Identifying Principal Components 60
   3.10 A Diagnostic Method 63
   3.11 A Generative Model 64
      3.11.1 Formulating the Model 64
      3.11.2 Fitting the Model 66
      3.11.3 Testing Validity of the Model 70
   3.12 Perception and Behavior 72
   3.13 Discussion 74

4 Discussion and Conclusion 77
   4.1 Connecting Two Forms of Behavior 77
   4.2 Limitations 79
   4.3 Directions for Future Work 80
      4.3.1 Flies 80
      4.3.2 ASD 82

Appendix A Additional Drosophila graphs 91

Appendix B Additional ASD graphs 95
## List of Figures

2.1 Sample fly path with visible cones ........................................... 19  
2.2 Sample fly path with invisible cones ...................................... 20  
2.3 Directional velocity of a segment of fly’s path .......................... 21  
2.4 Angular velocity of a segment of fly’s path .............................. 22  
2.5 Curvature of a segment of fly’s path ....................................... 23  
2.6 Sample segment of path in which fly interacts with only one obstacle 24  
2.7 Scatter plots of speed, angular velocity, and curvature versus distance from obstacle ................................................................. 26  
2.8 Softmax activation for the *Drosophila* model ............................ 30  
2.9 Simulated fly paths for one visible obstacle .............................. 32  
2.10 Scatter plots of speed, angular velocity, and curvature versus distance from obstacle for simulated data ................................................................. 34  
3.1 Template for the nine shapes traced ....................................... 42  
3.2 Sample neurotypical and ASD trials for the nine shapes ............... 42  
3.3 Example of segmenting the trial into individual repetitions ........... 43  
3.4 Template and sample curvature trajectories .............................. 45  
3.5 Example of how warping functions change a trajectory ............... 47  
3.6 An example of aligned curvature trajectories ............................ 50  
3.7 An example of aligned SRSFs ............................................... 50  
3.8 An example of warping functions ............................................. 51  
3.9 Histograms of maximum curvature for each shape .................... 52  
3.10 Histograms of variance in amplitude across the nine shapes ......... 54  
3.11 Karcher mean trajectories for each of the nine shapes ............... 56  
3.12 Histogram of speed for the neurotypical group versus the ASD group 58  
3.13 Scatter plot of variance in amplitudes versus speed across all shapes 59  
3.14 Scatter plot of variance in amplitudes versus speed across shapes 1 through 3 ................................................................. 60  
3.15 Percentage of variance explained by each principal component in shapes 3 and 4 ................................................................. 61  
3.16 Histogram of projections of curvature trajectory on each of the first three principal components for shapes 3 and 4 ......................... 62  
3.17 First three principal component trajectories for shape 4 ............ 65  
3.18 Example of smoothing curvature trajectories ........................... 66  
3.19 Illustration of how the *c* parameter affects the dynamics of the model in Equation 3.1 ................................................................. 68  
3.20 Examples of different fittings .............................................. 68  
3.21 Boxplot of *τ* values for the neurotypical and ASD groups ......... 69
3.22 Example of adding noise to simulated trajectories .................................. 70
3.23 Karcher mean trajectory of simulated data ............................................. 71
3.24 Variance in amplitudes for simulated data .............................................. 72
3.25 Scatter plot of perception task versus variance in amplitudes .................. 73

A.1 Additional simulated paths ................................................................. 92
A.2 Ratio of deterministic term to error term for position and heading based on short or long distance away from obstacle ........................................... 93

B.1 Demonstration of the $\frac{2}{3}$ power law .................................................. 96
B.2 Scatter plot of variance in amplitudes versus speed across shapes 4 through 6 97
B.3 Scatter plot of variance in amplitudes versus speed across shapes 7 through 9 98
B.4 Percentage of variance explained by each principal component in all shapes 99
B.5 Histogram of projections of curvature trajectory on each of the first three principal components for the first three shapes ........................................ 100
B.6 Histogram of projections of curvature trajectory on each of the first three principal components for shapes 4 through 6 ................................. 101
B.7 Histogram of projections of curvature trajectory on each of the first three principal components for shapes 7 through 9 ......................................... 102
B.8 $\tau$ values for shape 4, based on four separate speed/template settings .......... 103
B.9 $\tau$ values for shape 5, based on four separate speed/template settings .......... 104
B.10 Scatter plot of the block design task versus variance in amplitudes ........... 105
B.11 Scatter plot of the matrix reasoning task versus variance in amplitudes ....... 106
B.12 Scatter plot of the motion perception task versus variance in amplitudes .... 107
List of Tables

2.1 List of parameters and variables for the *Drosophila* model . . . . . . . . . . . 30
Chapter 1

Introduction

The phenomenon of movement has piqued the interest of the scientific community for centuries, and much remains to be discovered and understood. As the scientific community has discovered over time, just as each phenomenon has its unique properties and beautiful particularities, the underpinnings of these phenomena have been discovered to be remarkably similar. The two forms of movement we address in this thesis comprise only a small part of a gargantuan body of research, but the lessons learned from our analysis are widely applicable. In this thesis, we restrict ourselves to understanding just two specific forms of movement: *Drosophila* walking in a two-dimensional virtual reality environment with a grid of obstacles and people drawing simple shapes on an electronic tablet. In the former, we strive to understand how the flies’ movement is governed by their purpose in walking as well as their interaction with the grid of obstacles. In the latter, we employ this movement as a tool for differentiating between people with Autistic Spectrum Disorder (ASD) and the neurotypical population.
In both movement schemes, the subjects in question incorporate visual cues into their movement, thereby affecting their displayed motion profiles – specifically the positional and angular movement trajectories in the walking flies and the curvature and speed trajectories of the people in the tracing task. We aim to differentiate between two different modes of movement in both scenarios: the exploring and traveling behaviors in the flies and the tracing abilities of the ASD and typically developed populations. This analysis begins with gathering key characteristics in which we observe differences in the two modes of movement, which are then used to inform the formulation of a generative model for the movement in question. We can measure the validity of this model by generating a simulated data set, which undergoes the same initial analysis and hopefully displays the same differential characteristics and observations we identified. This not only informs our understanding of the movement, but it also helps us characterize the movements in question by identifying features of each mode of movement that likely drive the dynamics of the movement.
Chapter 2

Walking Behavior of *Drosophila melanogaster*

*Drosophila melanogaster* has played a huge role in the scientific community as the quintessential model organism. Many fascinating and important scientific procedures have been tested on flies due to their fast reproductive rate and relatively simple genetic code. These experiments on flies have led to experiments with more complex organisms, including humans, and the results of these studies have gone on to impact countless lives worldwide.

For this project, we take a step towards determining how flies interact with and respond to stimuli in their environment through observing and analyzing various aspects of their movement. Then, we formulate a generative model to mimic fly movement, incorporating key characteristics of this movement unearthed through numerical analysis. We show that our generative model does indeed display the essential traits of fly movement, in particular the two modes of movement, exploring and traveling, that we hypothesize govern the dynamics of movement in this environment.
CHAPTER 2. WALKING BEHAVIOR OF DROSOPHILA MELANOGASTER

2.1 Related Work

Studying the two-dimensional movement of *Drosophila melanogaster* has been a subject of scientific interest for quite some time. Tortorici and Bell (1988) conducted an experiment in which flies were placed inside an array of patches, among which resources were distributed of varying size and shape. The researchers then determined the degree of success the flies achieved in searching for and successfully finding these resources [1]. They looked at three types of flies: control, sitters (mostly stay in one place), and rovers (more likely to roam around); they found that the searching behavior exhibited by these flies was strongly demonstrative of their genotype. In fact, the control-line flies were most successful in locating resources, since their intermediate genotype allowed them to fluctuate between a local search phase to find resources nearby and a roaming phase to find resources elsewhere. They determined that, in general, flies frequently leave some areas unexplored because their movement takes on two modes: one is a rapid, nearly linear movement towards or away from something, and the other consists of multiple bouts of local search that prioritize revisiting previously explored spots over exploring a new spot.

Creamer et al. (2018) explored the behavior of *Drosophila* in response to seeing a visual stimulus in its environment [2]. The authors presented a visual rotational stimulus to the flies and measured both their turning response as well as the slowing of their walking speed as a result of this stimulus. However, when the flies were presented with translational stimuli, they did not display a rotational response but still displayed a decrease in walking speed. They also showed that the walking speed of flies is regulated by some mechanism that is dependent on the speed of visual motion, but not on the spatial arrangement of the visual stimuli nor the relative direction of the visual stimuli. This mechanism is characterized by the behavior of slowing down when a visual stimulus is nearby. To produce this effect, the
authors constructed an algorithm that is multi-layered, incorporating the speed of the visual stimulus as well as a mechanism of visual stabilization.

Robie et al. (2010) studied the behavioral response of a fly in response to a simple or topologically complex landscape with different sized cones arranged in an array [3]. Previous research has shown that *Drosophila* uses a combination of sight, smell, and touch to navigate them towards sources of nutrition. Furthermore, when presented with a prominent visual object, flies will turn towards and walk towards it, even after the object disappears [4,5]. The researchers found that the flies did show interest in visual objects, but this interest did not vary based on the shape of these visual objects. In fact, once they arrived at the visual object, they did show a preference for tall, steep objects as shown by a longer period spent exploring or resting on that object. They also found that the appearance of visual stimuli can affect the speed, walking duration, and angular velocity.

Many researchers have explored the mechanics and dynamics of the movement of flies and how they respond to visual and other stimuli in their environments. An up-and-coming field in this sphere is using virtual reality (VR) as a tool to better analyze this movement. VR environments have been shown to be effective at mimicking a real arena, producing the same results that have been achieved in a real-life scenario [6].

### 2.2 The Data

Our data comes from the Jayaraman group at the Howard Hughes Medical Institute’s (HHMI) Janelia Research Campus, where Hannah Haberkern is conducting some experimental work with *Drosophila melanogaster* in a two-dimensional virtual reality (VR) environment [6]. In this environment, flies are head-fixed and are walking on a spherical
treadmill, which tracks their two-dimensional movement. The fly is surrounded by a screen that displays the virtual environment the fly is walking in.

To mimic real movement in a two-dimensional environment, Haberkern et al. developed a program in C++, FlyOver, that translates these in-place movement measurements into a real two-dimensional movement. This program was used to dynamically update the view of the fly as it “moved around” in its environment. The environment had a grid of obstacles in the form of cones and cylinders that the fly could walk to, away from, or even circle around. In the environment, the fly’s visual distance was controlled with a “virtual fog” that would reveal an object when it passed a threshold of distance from the fly. In addition to being able to interact with the environment, the flies were optogenetically stimulated in some trials with either a sweet taste that would attract it towards an obstacle, or a feeling of heat that would turn it away from an obstacle. Also, some trials were conducted with invisible obstacles; that is, flies would be optogenetically stimulated around an invisible obstacle. Haberkern et al. found that this two-dimensional virtual reality environment was conducive to reproducing some aspects of fly behavior that had previously been demonstrated in a free-walking fly setup, namely a local search behavior triggered by the optogenetic stimulation of the sweet taste as well as a learned avoidance of optogenetically induced aversive heat near certain obstacles.

Figures 2.1 and 2.2 show the path of a fly walking in a single trial with visible cones. We can see many different signatures of movement in these two graphs. We focus on where the fly seems to stay in one area and explore, versus where the fly seems to be traveling in a relatively straight line to another destination. For the portions of the path that are inside the hexagon of obstacles, most of the exploring behavior seems to occur either around an obstacle or fluctuating among a small group of obstacles. For instance, in the center-right of
2.2. THE DATA

Figure 2.1: Sample fly path with visible cones. An example of a path that explores a large portion of the arena and spends relatively more time in the traveling mode of movement.

Figure 2.1, the fly seems to be walking among four or five obstacles in an extremely erratic fashion. We can see the same phenomenon in the upper left of Figure 2.2, in which the fly in this trial seems to spend most of its time exploring its near surroundings and less time traveling to obstacles in other locations. By contrast, the fly in Figure 2.1 seems to be traveling among three large clumps of obstacles, spending quite some time exploring within these clumps before traveling in a relatively straight line to another area to explore. We can distill these erratic motions into two types of motion.

The first is the exploring behavior that is characterized, informally, by a twisty and winding path that seems to stay in one area. Mathematically, the angular heading of the fly changes frequently, and these changes seem to be large in magnitude. This behavior can be observed around a certain obstacle, indicating that a fly is exploring an object by walking...
Figure 2.2: Sample fly path with visible cones. An example of a path that explores a smaller portion of the arena and spends relatively more time in the exploring mode of movement.

around it, or within a clump of obstacles, which could represent a level of uncertainty about which obstacle to explore.

The second is the traveling behavior that is marked by a relatively straight and less windy path that appears to move substantially in some direction. In other words, the heading of the fly does not seem to change as much as it walks on its almost straight path towards another destination. One interpretation of this behavior is that the fly could become uninterested in its current location and start moving towards another location, either because of some attractive feature of the destination or some repulsive feature of its current location.

Our goal is to formulate a generative model that emulates our findings from the numerical analysis in the following section.
2.3 Initial Analysis

We commence our numerical analysis by looking at the positional velocity, angular velocity, curvature, and tangent vector for the paths. Because each trial is rather long, we focused on a 2-second long section of the trial.

The first quantity we examined was the positional velocity. In Figure 2.3, the fly stops moving for two substantial periods of time, highlighted in the black circles. Before the first stopping, the fly seems to be moving in a relatively straight path, which is illustrated by $\frac{dx}{dt}$ being almost entirely positive and $\frac{dy}{dt}$ being almost entirely negative. For the first stopping (on the left), the fly seems to move in a small loop before continuing on in its path, and its movements are small in magnitude right before stopping as indicated by the velocity graph. The fly then loops around using slightly erratic movements and a slower speed than we saw at the beginning of this segment of the path. Once it reaches near the bottom of the loop, however, we see a return to the behavior at the beginning with both $\frac{dx}{dt}$ and $\frac{dy}{dt}$ positive for a
significant amount of time. Then, the fly seems to completely stop at the second black circle before resuming its previous speed, which indicates that the traveling mode of movement consists of more than just a high walking speed with little angular change.

The natural next step is to examine the angular velocity of the fly, as in Figure 2.4. The first section of the path before the first stopping has little to no angular change, as expected, since it seems to be almost a straight path to the first black circle. However, at the first stopping, we observe a spike in angular velocity as it turns around in a clockwise direction, but stays in place as it does so. As it curves around, the angular velocity is relatively large in magnitude until it comes to the second stopping, at which there is almost no angular change, before it goes in an almost straight line to the end of the segment.

Next, we can examine the curvature of the path, as shown in Figure 2.5. To compute curvature, we need to ensure continuity of the first and second derivatives, so we performed a cubic spline interpolation of the path, which is shown in magenta in the graph on the left.
2.3. INITIAL ANALYSIS

Figure 2.5: Left: segment of fly’s path with key areas of high curvature highlighted with black circles. Right: curvature trajectory over the segment on the left. The same areas of high curvature are highlighted with black circles.

This spline interpolation has less granularity than the original data so we can better see the curvature trajectory in the graph on the right side. The peaks in the curvature do not necessarily correspond to the features of the path that we examined before. In this case, the peaks in curvature occur at places in the path where the fly traces a small semicircle-like dip in its path, as we see for the first two black circles. The third black circle occurs where the fly comes to a stop while it is moving in an almost straight line, and the fly seems to curve around a bit before continuing on. As expected, the first part of the path has a low curvature, as does the last part of the path (excluding the third black circle). As the fly widens on its spiral around after the first peak, the curvature gradually decreases.
2.4 Identifying Two Modes of Movement

A natural question to ask after analyzing the dynamics of the flies’ movements is to see whether they change based on distance from obstacles. According to our earlier theory, we distinguished between two behaviors: exploring and traveling. We would expect that the traveling behavior occurs more when a fly is far away from an obstacle, so both the angular velocity and curvature would be low, while the positional velocity would be relatively high, since the fly is traveling in more or less of a straight line. By contrast, we would expect the exploring behavior to occur more near obstacles, where the angular velocity and curvature are relatively high, while the positional velocity would be relatively low.

To this end, we chose segments of paths from ten flies where the path was relatively straight except near obstacles, at which the fly circled around or was more erratic in its movement. An example of such a path is shown in Figure 2.6. Choosing these paths was relatively hard to do, since the fly does not simply interact with one obstacle: it is, after all, on a grid, so there are likely many obstacles at any point in time that it can see and
2.4. IDENTIFYING TWO MODES OF MOVEMENT

decide to interact with. However, to form a basis for a generative model of movement, we
must analyze a fly’s interaction with just one obstacle. Therefore, when choosing paths to
use for this analysis, we tried to choose segments that appeared to only be affected by one
obstacle, as the one in Figure 2.6. Of course, there is no way of knowing what exactly the
fly was thinking when it navigated through these cones, but this is a good approximation
of isolation to just one obstacle. With this path, we see clearly that there are many more
slowdowns (blue points that represent lower velocity) near that obstacle and higher velocities
(in green) away from the obstacle. Interestingly, the higher velocity points near the obstacle
seem to occur when the fly is circling around the obstacle. One last note is that even in the
portions of the path that are far from the obstacle, there are moments of low velocities and
slowdowns, perhaps due to the other obstacles that are in sight. We saw a similar effect in
Figure 2.3.

Aligning with the theories about how positional velocity, angular velocity, and curva-
ture change with distance from said obstacle, we then analyzed how these three quantities
changed with distance from the obstacle. Although there were many obstacles around, dis-
tance from the obstacle was calculated only for the one obstacle that each fly interacted
with, so we could simulate something close to a one-obstacle environment. Furthermore,
due to the high levels of noise in the dataset and to declutter each scatter plot, we removed
the outliers by separating the distances into bins of five centimeters each and removing the
top and bottom ten percent. Even with these modifications, the noise in the data is still
palpable, but this is perhaps unavoidable when using multiple flies in multiple trials that
perhaps have different flavors of movement.
CHAPTER 2. WALKING BEHAVIOR OF DROSOPHILA MELANOGASTER

Figure 2.7: Top: scatter plot of speed versus distance from obstacle that displays a positive correlation and the “T” shape. Left: scatter plot of angular velocity versus distance from obstacle that displays a negative correlation and the “L” shape. Right: scatter plot of curvature versus distance from obstacle that displays a negative correlation and the “L” shape.
2.4. IDENTIFYING TWO MODES OF MOVEMENT

Figure 2.7 shows scatter plots and the least-squares regression line through these points for each of the three quantities: speed, angular velocity, and curvature. We took the amplitude of the positional velocity vector to calculate the speed.

Looking first at the relationship between speed and distance from the obstacle, from our previous discussions, we would expect that the fly would slow down closer to the obstacle and speed up farther away from the obstacle. Thus, the two would be positively correlated, which is exactly the effect seen in Figure 2.7. An interesting part of this graph is the cluster of points in a vertical segment to the left of the graph, at a distance of close to zero. This physically translates to a wide variety of speeds at small distances from the obstacle, potentially due to something similar to what we observed in Figure 2.6, where the fly circled around the obstacle at a moderate speed while also hovering near the obstacle in place without circling around. Another feature of interest is that the graph is not particularly linear, but in more of a “T” shape, where the distribution of positional velocities is very wide for small distances but much smaller for larger distances. In fact, taking out the cluster of low-velocity points at very close distances would cause the least-squares line to become almost horizontal. This abrupt change in the structure of the graph hints that the two modes of movement (traveling and exploring) could be almost disjoint: the switch between one and the other is perhaps rather sudden.

With regards to angular velocity, we would expect that a closer distance to the obstacle would yield a greater angular velocity due to the exploratory nature of this movement, whereas a greater distance would yield a smaller angular velocity. In Figure 2.7, we do see a slight negative correlation between the two quantities, though there are a few features of interest in this graph. Firstly, we see another vertical cluster of points below the least-squares line, which defies our earlier theory. However, it could be the case that some of
these points are part of circling around the obstacle, or various stops around the obstacle that happen to be of small angular change. There also seem to be a fair amount of low angular velocity data points throughout all ranges of distances, which can be somewhat expected as the fly moves in short but likely straight lines in various points of the path. As with the previous graph, the curve seems to take on an “L” shape, where there are quite a few points of high angular velocity very close to the obstacle, but much fewer of these occurrences at further distances. This again suggests that there may be a relatively sharp transition between the traveling and exploring phases.

Finally, looking at the relationship between curvature and distance from obstacles, we would again expect a negative correlation since the fly is likely making smaller loops closer to the obstacle and turns of a larger radius farther away. The graph in Figure 2.7 does exhibit this negative correlation and moreover looks similar to the angular velocity graph, perhaps because the two quantities are much related. Again, we observe many high curvature data points at extremely low distances, but fewer of these points at further distances, while there are quite a few low-curvature data points throughout the range of distances, suggesting the same “L” shape.

One note for all three of these graphs is that the distribution of points tends to skew towards points associated with lower distances from the obstacle. However, this is due to the nature of the data: because the flies slow down closer to objects and speed up farther away, the concentration of data points around the obstacles is higher than the relative proportion of points far away from the object. Nevertheless, our findings from the above analysis support our initial theories about the two modes of movement, as well as inform various characteristics of the generative model that will be discussed next, such as the weakly
positive correlation between speed and distance, but perhaps most importantly, that the switch between our hypothesized modes of movement should be more sudden than gradual.

2.5 A Generative Model

Keeping the previous sections’ numerical analysis in mind, we formulated the following generative model.

\[
\begin{align*}
\frac{dx}{dt} &= v(d) \cos \theta + \epsilon_x \\
\frac{dy}{dt} &= v(d) \sin \theta + \epsilon_y \\
\frac{d\theta}{dt} &= \left( k_{\epsilon_\theta} + \frac{k_{\epsilon_\theta}}{1 + e^{d-5}} \right) \epsilon_\theta - \frac{1}{1 + e^{-k_\theta(d-5)}} \sin(\theta - \theta_0)
\end{align*}
\] (2.1)

Equation 2.1 captures the generative model, with variables and distributions defined in Table 2.1. This equation specifically only governs a scenario with a single fly in an environment with only one object, located at \((0,0)\) for simplicity. Our model dictates how \(x, y\), and the heading of the fly changes over time. The equations are interdependent and each has a noise term as well as a deterministic term. To govern the position of the fly, we began with a simple equation, \(x'(t) = v(t) \cos \theta\), and added a dependence on \(d\), the distance between the fly and the obstacle, to capture our observation in Figure 2.7 that speed increased with distance away from the obstacle.

This dependence on \(d\) is a variation of a softmax activation function, an example of which is shown in Figure 2.8, where we set \(k_\epsilon = 1\). \(k_\epsilon\) controls how sharp the slope is from
CHAPTER 2. WALKING BEHAVIOR OF DROSOPHILA MELANOGASTER

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$\sqrt{x^2 + y^2}$</td>
</tr>
<tr>
<td>$\epsilon_x, \epsilon_y$</td>
<td>$\sim \mathcal{N}(0, \sigma^2)$</td>
</tr>
<tr>
<td>$\epsilon_{\theta}$</td>
<td>$\sim \text{Unif}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$</td>
</tr>
<tr>
<td>$\theta_0$</td>
<td>heading of fly relative to obstacle</td>
</tr>
</tbody>
</table>
| $v(d)$ | $\begin{cases} 
1 \\
\frac{1}{1 + e^{-k_v(d-d_0)}} \cdot \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}
\end{cases}$ with probability $p$
| $\frac{1}{1 + e^{-k_v(d-d_0)}} \cdot \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$ with probability $1 - p$
| $k_v, k_{\theta}, k_{\epsilon}, \sigma, p$ | tuneable parameters |

Table 2.1: List of parameters and variables for the Drosophila model in Equation 2.1

Figure 2.8: Illustration of a classic softmax activation function, modified versions of which are used in Equation 2.1.
2.5. A GENERATIVE MODEL

0 to 1, and the softmax was chosen because it could provide a relatively sharp transition from one mode of movement to the other, as observed in the previous section. Moreover, the sharpness of this transition could be directly controlled by tuning $k_v$. We observed that after some period of time had passed, the fly would leave its location and move on to another destination, so we added a probabilistic increase in velocity. Without the random chance that the velocity increases regardless of the fly’s current position and velocity, the fly would just eventually come to a stop. The noise terms $\epsilon_x$ and $\epsilon_y$ for $\frac{dx}{dt}$ and $\frac{dy}{dt}$ were chosen to be Gaussian, where the variance of the Gaussian was ultimately chosen to be 0.1.

For the change in angular heading, we started with a noise term that is a Uniform distribution on $(-\frac{\pi}{2}, \frac{\pi}{2})$, since we saw a very wide distribution in angular change. Then, to incorporate the observation that this distribution was wider at shorter distances than at wider distances, we multiplied this noise by a modified softmax that has a value of 1 close to a distance of zero, and tapers down to $k_{\theta} \in (0, 1)$ for larger distances. $k_{\epsilon_{\theta}}$ was ultimately chosen to be 0.7.

Finally, the second term is a deterministic term: we began with $-\sin(\theta - \theta_0)$, which would turn the fly back towards the obstacle if the fly was turned away. The multiplier is a softmax function, which has the effect that the deterministic term dominates if the fly is far away, and the noise dominates if the fly is close to the obstacle.

Figure 2.9 shows some of the simulated paths, with more in Appendix Figure A.1. The variability in the paths mirrors the noise we observed in the data itself, and the speed distribution matches well to what we observed. Around the obstacle, the fly mostly slows down (blue areas), and the fly speeds up away from the obstacle (green areas). However, as we saw in the segment in Figure 2.3, the fly sometimes slows down even when it is in the traveling mode of movement. We also see a variety of paths to the obstacle: some are
Figure 2.9: Simulated fly paths for one visible obstacle. Upper left: the fly takes a rather windy route to the obstacle, then slows down considerably but does not change much in angular velocity as it passes by the obstacle. Upper right: the fly takes a more direct route to the obstacle, then slows down and appears to explore the obstacle briefly before leaving. Lower left: the fly takes a more windy route to the obstacle and moves in a slow loop around the obstacle before walking off. Lower right: the fly takes a more direct route to the obstacle then slowly walks around it before walking away. It also displays a slowdown as it walks away.
2.6. DISCUSSION

straighter than others, but we do see a general trend of noisy movement that eventually does reach the obstacle.

In Figure 2.10, we replicated the analysis done in Figure 2.7 but for the simulated paths. Although the similarity may not be immediately visible at first, the key elements that we discussed still appear. The speed graph has the same upward slope, and a sort of “T” shape as discussed, while the angular velocity and curvature graphs both have negative correlations and “L” shapes. However, the slope of the angular velocity plot is not very steep, which could indicate that the model is missing another piece of information that could improve its similarity to what we observe in real life. Furthermore, the numerous tuning parameters could be adjusted to mimic the results obtained from the data; for instance, the width of the switch between the two modes of movement could be shortened to match the smaller width of the vertical portions of the “L” graphs. Additional analysis of the ratios of the deterministic terms to the noise terms can be found in Appendix Figure A.2.

2.6 Discussion

In this chapter we analyzed fly paths in a two-dimensional virtual reality scheme and identified key features of interest for the geometry and dynamics of the flies’ movement. Previous research has found that flies switch between two modes of movement: a local search mode in which flies slow down in the presence of some stimulus (visual, olfactory, or gustatory), and an exploitative mode of movement that causes it to travel from one place to another in its search for these resources. We observed this distinction between the modes of movement in our analysis of the fly walking data: we found a positive correlation between speed and distance from the object, but upon closer examination, the positive correlation was not linear; in fact, it consisted of a wide range of speeds at a small distance to the
Figure 2.10: Left: scatter plot of speed versus distance from obstacle for simulated data that displays a positive correlation and the “T” shape. Right: scatter plot of angular velocity versus distance from obstacle for simulated data that displays a negative correlation and the “L” shape. Center: scatter plot of curvature versus distance from obstacle for simulated data that displays a negative correlation and the “L” shape. All three of these graphs display the same overall shape as those in Figure 2.7.
obstacle and a narrower but larger in magnitude distribution of speeds for greater distances from the obstacle (the “T” effect). We found that angular velocity had a negative correlation with distance from the obstacle, but also discovered that this relationship was not linear. Close to the obstacle, the angular velocity had a wide range of values as the fly performed this local search movement, but farther away from the obstacle the angular velocity was much closer to zero, and the width of this distribution appeared to be approximately the same for most of the larger distances (the “L” effect). Very similar results were obtained for the relationship between curvature and distance from the obstacle: an overall negative correlation, driven by the wide range of values at a short distance and low values of curvature at greater distances.

This analysis informed our conception of a generative model for two-dimensional movement. The key element was the observed switch-off between the exploratory (local search behavior) and the traveling (exploitative search behavior) of the fly, which was achieved through use of modified softmax functions that would switch “on” and “off” at a small distance away from the obstacle. This softmax function was applied to both the velocity (decreasing at closer distances to the obstacle) as well as the mechanism for attracting the fly back towards the obstacle, which turned the fly towards the obstacle if it was turning away through a sinusoidal method of control. Finally, the noise term on the angular change equation was carefully calibrated using the softmax function to increase when the fly is close to the obstacle and decrease when the fly is further away. Analysis of the simulated paths generated from this model revealed that it had the characteristics that we found to be features of interest from our initial examination of the fly data. This indicates that our generative model is a good approximation of how flies interact with an obstacle in a two-dimensional walking environment.
Chapter 3

The Motion of Tracing in People with ASD

Autistic Spectrum Disorder (ASD) is defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) as being characterized by persistent deficits in social interaction and/or communication as well as repetitive and restrictive behavioral patterns. Previous studies have demonstrated motor differences in children with ASD, such as macrographia (large handwriting) and impairment in manual dexterity.

For this project, we seek to extract meaningful features from a tracing task completed by two groups of ASD and neurotypical (NT) subjects. In particular, we examine the curvature of the shapes traced and analyze differences around areas of high or low curvature, as well as differences that stem from speed or variance of shapes traced. From these extracted features, we formulate a generative model that aims to capture the tracking delay of ASD individuals and identify key parameters that govern the dynamics of this model. Then, we use this model to generate a simulated dataset and show that this dataset has the same
38 \textit{CHAPTER 3. THE MOTION OF TRACING IN PEOPLE WITH ASD}

features that were identified as differentiating between the two groups by reproducing the features we found through numerical analysis using simulated data.

3.1 Related Work

The body of research surrounding ASD has brought to light numerous findings that have expedited early diagnosis, as well as the formulation of various methods of education and guidance during a person’s younger years. The range of behavioral and social markers have been well studied and documented, but the causes and potential treatments for people with ASD are mostly unknown and require further research and investigation. The diagnosis process of ASD is extremely limited: it largely relies on studying how the individual develops linguistic skills as well as observing the individual in social settings. Recently, the eyesight of people on the autistic spectrum has been a growing subject of major analysis, particularly their sensory perception as well as visual symptoms.

Ashwin et al. (2009) found that people with ASD had an extremely high level of visual acuity, even though other studies have shown that people on the autistic spectrum report more refractive errors in their vision \cite{9}. Furthermore, studies have not found that people with ASD display a significantly lower threshold for contrast sensitivity \cite{10}. However, there are conflicting studies regarding how the ASD population performs in shape discrimination (i.e. distinguishing between squares and rectangles) in that De Jonge et al. (2007) found no disparity in the ASD population whereas Ropar and Mitchell (2002) found that the ASD population performed worse than their neurotypical counterparts \cite{11,12}. People with ASD have also been shown to have more difficulty integrating information from different sources to complete a task, but Happe (1996) demonstrated that the ASD population were less susceptible to visual illusions than the neurotypical population \cite{10,13}. The variety in these
results demonstrate that the literature on visual performance of the ASD population is still under review and that much more research and documentation is needed to tease out the mechanisms underlying of these observed behaviors.

Stevenson et al. (2017) found that for the ASD population, greater attention to detail was correlated to a decrease in temporal adaptation [14]. They devised a task in which participants were first adapted to relationships between pairs of consistent temporal stimuli and then performed a judgment task in which the same pairs were temporally offset. The participants were then measured on their temporal adaptability, and they found that this measure was related to the participant’s Attention to Detail submeasure within the Autism Quotient. This submeasure focuses on the individual’s tendency to focus on small details, while missing the larger, overarching structure. Thus, the study found that the higher the participant scored on the Attention to Detail submeasure, the less the participant was able to adapt their visions of temporal synchrony to a temporal offset. This study also supports previous findings that people on the autistic spectrum had more trouble integrating prior information with current information. This impairment has been linked to various long term effects, particularly in language acquisition. However, Stevenson et al. did not find a link between the participants’ temporal integration abilities and their speech perception abilities.

The drawing abilities of the ASD population have been studied extensively as well, as drawing is a useful and relatively easy potential method of early ASD diagnosis. Jolley et al. (2013) examined the expressive (happy or sad) drawings of children with ASD and contrasted them with a neurotypical sample of children that matched the ASD sample in either mental age or learning difficulty [15]. On an expressivity scale, the two populations actually performed very similarly and drew the same numbers of figures, but the researchers
observed that the ASD drawings of specific figures contained more immature forms. Specifically, their depictions of happiness and sadness appeared in a literal fashion (i.e., drawing happy or sad faces on people) rather than being demonstrated by a certain social scene (i.e. a celebration to depict happiness). Moreover, the ASD sample drew fewer social scenes than their neurotypical counterparts, indicating that the well-documented social impairment of ASD individuals also appeared in their expressive drawings.

Van Ommeren et al. (2015) conducted another drawing study, this time measuring the perception of reciprocity in the ASD population [16]. Studies have shown that the ASD population lacks in its ability to identify and participate in reciprocation compared to their neurotypical counterparts. In fact, the DSM-5 includes poor reciprocity as a necessary criterion for diagnosis, increased from its appearance as a possible criterion in the DSM-4. Van Ommeren et al. (2015) used the Interactive Drawing Test (IDT), which involves the researcher and participant “drawing together” (i.e., researcher draws a line and pushes the paper to the participant, alternating turns to add to the drawing). Various elements were added by the researcher (interfering, absurd, and damaging), and the participant either accepted or rejected. They found that the ASD group took fewer turns and displayed less reciprocal turn taking. Furthermore, the ASD group did not collaborate on a mutual subject as often as the neurotypical group and contributed less to an object initiated by the researcher. Finally, the ASD group did not accept the researcher’s contributions as frequently as the neurotypical group.

The vast body of research on ASD individuals indicates that the symptoms are wide-ranging and much more research is needed to distill the key mechanisms underlying the disorder. Nevertheless, the spectrum of behaviors observed can inform analysis of their underlying mechanisms and better connect the pieces of our understanding of this disorder.
3.2 Tracing Data

Two groups (ASD and NT) of ten subjects completed various cognitive tasks as well as a tracing task. Each group consisted of ten subjects, all between the ages of 24 and 40 with varied genders and handedness. There were eight male and two female subjects in each group, as well as eight right-handed and two left-handed subjects in each group.

For each subject, tracing data was gathered over four meetings. In each session, the subject drew nine shapes, randomly ordered. The subject would observe the template for three seconds, then the template would either remain while the subject traced the shapes or it would disappear as the subjects did the tracing from memory. Two meetings were done with the template and two additional meetings were done from memory. Within each meeting, there were four sessions that differed in the speed at which the subject was instructed to draw the shapes. The first session was at the participant’s own pace, the second and third sessions were at slow and fast paces (randomly ordered), and the last session was at an intermediate pace, defined as in between the slow and fast speeds. For each session, the subjects drew nine shapes ordered at random. For each shape, the drawing time was 37 seconds, and the sampling rate of the tablet was 146 Hz.

The templates for the nine shapes are shown in Figure 3.1, and sample trials for an ASD subject and a neurotypical subject are shown in Figure 3.2.

Since each trial was standardized at 37 seconds each, the number of complete tracings varied by both the shape and the speed at which each shape was traced. For better analysis, the data from each trial was segmented into individual tracings of the shapes. An example of segmentation for shape 1 is shown in Figure 3.3. The segmentation was standardized such that each tracing started at the bottommost point of the shape. As a heuristic, a graph of the $y$-coordinates of the tracing was created, and the segmentations generally were cut off.
CHAPTER 3. THE MOTION OF TRACING IN PEOPLE WITH ASD

Figure 3.1: Template for the nine shapes traced.

Figure 3.2: Sample neurotypical (left) and ASD (right) trials for each of the nine shapes from two individuals. All shapes drawn were at the participant’s own speed and with the template. The same individual is depicted for all nine shapes in each graph.
Figure 3.3: Left: sample segmentations of a particular trial into nine individual repetitions with start and stopping points. Right: $y$-coordinate of a trial, with blue circles indicating where the segmentations would start and end.
at the lowest points in these graphs, as can be seen in Figure 3.3 for shape 1. This was calculated using the \texttt{findpeaks} function in MATLAB.

Due to the chaotic nature of the movement during the beginning and end of each trial, a few repetitions done at the beginning and the end of the trial were thrown out, as were trials where the participant made a mistake. For instance, in shape 1 (a loop-de-loop), if the subject accidentally drew two large loops in succession instead of alternating large and small loops, this half-repetition would be thrown out. Visually, this would appear as two successive dips to about $-7.5$ in the $y$-coordinate, and we would remove this by ignoring the portion of the trajectory in between these two successive dips.

### 3.3 Curvature Calculation

Due to the large variability within the tracing of the data, it would be difficult to analyze the traced shapes themselves. One transformation that reduces the variability in the data is the curvature of the tracings, especially the magnitude of the curvature as well as how quickly the curvature changed in different areas of each shape.

Thus, we calculated the curvature of each tracing at various time points along the curve, standardized to a sampling rate of 200 points per tracing. Furthermore, since each of the eight speed/template settings was completed twice, the two identical trials were concatenated to yield 72 levels per individual (eight speed/template settings and nine shapes). For a parameterized curve $(x(t), y(t))$, the curvature at time $t$ is given by

$$
\kappa(t) = \frac{|x'(t)y''(t) - x''(t)y'(t)|}{\left((x'(t))^2 + (y'(t))^2\right)^{3/2}}
$$
3.3. CURVATURE CALCULATION

Figure 3.4: Template curvature trajectories for all nine shapes in black with sample curvature trajectories overlaid in red. The same neurotypical individual was used for all sample curvature trajectories, drawn at the individual’s own speed and with the template.
We performed cubic spline interpolation on both $x(t)$ and $y(t)$ to ensure continuous second derivatives as required by the formula above. Each tracing was normalized to 200 time steps between $t = 0$ and $t = 1$. The curvature trajectories for the template shapes as well as examples from the data are shown in Figure 3.4. For the first three shapes (the loop-de-loops), we observe an increase in peak curvature as the inner loop becomes smaller and smaller (and therefore more angular). For the next three shapes (the ovals), we observe the highest peaks in curvature in shape 4 because it is the farthest from a circle. The curvatures in shapes 5 and 6 are lower and less dramatic in terms of changing from low to high or high to low curvatures than the first three shapes. Finally, the figure eights are chaotic in their trajectories, having many small peaks in addition to multiple large peaks. However, note that all curvatures are less than 1, which is significantly less than shapes 3 and 4, for instance. Furthermore, these three shapes are unique in that they require the participant to change curving direction while drawing the shape.

### 3.4 Curvature Alignment Theory

In order to compare these curvature trajectories across participants and speed/template settings, we must first align the curvature trajectories, since for each trial, the peak of curvature will not occur at the same point. These methods are adapted from *Functional and Shape Data Analysis* by Srivastava and Klassen [17]. For instance, if an oval happens to be tilted, the peak curvature will occur earlier or later than the midpoint of the time scale. So, to calculate a mean trajectory or variances of trajectories, we will align the curvatures involved in each calculation. To do this alignment, we first start with a few definitions.
3.4. CURVATURE ALIGNMENT THEORY

Figure 3.5: Left: a function $f_0$ to be warped. Center: various warping functions $\gamma_i$ shown in different colors. Left: warping functions applied to $f_0$ to result in $f_0 \circ \gamma_i$. The warping functions primarily changed the lowest point of $f_0$. Taken from Functional and Shape Data Analysis by Srivastava and Klassen [17].

**Definition 1** *(Absolutely continuous).* A function $f : [0, 1] \to \mathbb{R}$ is an absolutely continuous function if it is differentiable almost everywhere on its domain and if for any $t \in [0, 1]$, we have $f(t) = f(0) + \int_0^t f'(u) \, du$.

**Definition 2** *(Square root slope function (SRSF)).* Given an absolutely continuous function $f : [0, 1] \to \mathbb{R}$, we define the square-root slope function (SRSF), $q(t)$, as

$$q(t) = \text{sign}(f'(t)) \sqrt{|f'(t)|}$$

Given an absolutely continuous function $f(t)$ defined on $t \in [0, 1]$, we can warp $f(t)$ using a function $\gamma : [0, 1] \to [0, 1]$ such that $\gamma(0) = 0$, $\gamma(1) = 1$, and $\gamma$ is strictly increasing. The warped function will then be $(f \circ \gamma)(t)$. For the warped function $f \circ \gamma$, the SRSF becomes

$$\tilde{q}(t) = \text{sign} \left( \frac{d}{dt} (f \circ \gamma)(t) \right) \sqrt{\left| \frac{d}{dt} (f \circ \gamma)(t) \right|}$$

For instance, Figure 3.5 depicts a function on the left, various warping functions $\gamma$ in the center, and the warped functions $f \circ \gamma$ on the right [17].
Warping the figures in practice requires optimization: suppose we have functions \( f_1, \ldots, f_n \) with corresponding SRSFs \( q_1, \ldots, q_n \). Then, we must choose an optimal set of warping functions \( \gamma_1, \ldots, \gamma_n : [0, 1] \to [0, 1] \) with the restriction that each \( \gamma_i \) is nondecreasing. This is done by iteratively improving the \( \gamma_i \) to better fit the warping objective.

First, we can consider the case of warping a function \( f_2 \) to align with a function \( f_1 \), with corresponding SRSFs \( q_1 \) and \( q_2 \). We start with an \( m \times m \) grid on \([0, 1] \times [0, 1]\) with gridpoints \((i/m, j/m)\) for \(0 \leq i, j \leq m\) and aim for a solution that is piecewise linear from \((0, 0)\) to \((1, 1)\) passing through points in this grid. For each point \((i/m, j/m)\) in the grid, we let \( \mathcal{N}_{ij} \) denote the neighbors of that point that are allowed to reach that point in this path. Due to our restrictions on the warping function \( \gamma \), the set of neighbors only consists of those to the left and below the grid point. We then can define a “cost” \( C_{ij,kl} \) of reaching point \((i/m, j/m)\) from a point in the neighboring set \((k/m, l/m)\):

\[
C_{ij,kl} = \int_{k/m}^{i/m} \left( q_1(t) - q_2(\gamma_{ij,kl}(t)) \sqrt{\gamma_{ij,kl}'(t)} \right)^2 dt,
\]

where \( \gamma_{ij,kl} \) is the straight line joining \((k/m, l/m)\) and \((i/m, j/m)\). Then, using this definition, we define \( D_{ij} \) as the optimal cost of reaching node \((i/m, j/m)\) from \((0, 0)\), so our ultimate goal is to solve for \( D_{mm} \). We define and calculate \( D_{ij} \) using a recursive approach:

\[
D_{ij} = \min_{(k/m, l/m) \in \mathcal{N}_{ij}} \left( D_{kl} + \int_{k/m}^{i/m} \left( q_1(t) - q_2(\gamma_{ij,kl}(t)) \right)^2 dt \right) \sqrt{\gamma_{ij,kl}'(t)}.
\]

With this definition, we can create an algorithm to compute the optimal warping function \( \gamma^* \):

1. Compute \( \mathcal{N}_{ij} \) for each \( i, j \).
3.5. **CURVATURE ALIGNMENT OF THE DATA**

2. Initialize $D_{00} = 0$ and $D_{i0} = D_{0j} = \infty$ for $1 \leq i, j \leq m$, since we do not allow our path from $(0,0)$ to pass through these infinity-cost points.

3. Compute $D_{ij}$ for the entire grid using a dynamic programming approach, using numerical approximations for the integral in the definition of $D_{ij}$. We can let $\hat{n}_{ij}$ denote the optimal $(k/m, l/m) \in \mathcal{N}_{ij}$ as computed for each point $(i/m, j/m)$.

4. To compute the optimal warping function $\gamma^*$, we use $(i/m, j/m) = (1, 1)$. The first segment of the line (going backwards) is from $(1, 1)$ to $\hat{n}_{ij}$. Then, setting our starting point as $\hat{n}_{ij}$, we repeat the algorithm until we obtain a starting point of $(0,0)$. This set of segments then forms our piecewise linear function $\gamma^*$ from $(0,0)$ to $(1,1)$.

Now, we must tackle the problem of fitting many $\gamma$ at once. However, this is somewhat easily achieved by running the above algorithm to align $f_2$ with $f_1$ first, then $f_3$ with $f_1$, and so on until $f_n$ is aligned with $f_1$ (as well as the others).

### 3.5 Curvature Alignment of the Data

In practice, the template curvature function is always set to be $f_1$ in our case, so we align all of the data trajectories to the template curvature trajectory.

We can see sample warping results in Figure 3.6. By setting the template trajectory as $f_1$, we can guarantee that the peaks are symmetrically located. Crucially, the peaks of the curvature trajectories are now all aligned so that we can perform various mean and variance calculations to generate summary statistics. There is also a wide variation in the heights of the trajectories in the areas of highest curvature, which we will analyze in the next section. Along with aligning the peaks, we observe that the dips in the curvature trajectories are also aligned and there is less noise in the low points of the aligned functions as well.
CHAPTER 3. THE MOTION OF TRACING IN PEOPLE WITH ASD

Figure 3.6: Aligned curvature trajectories in blue with the template overlaid in red.

Figure 3.7: Aligned SRSF trajectories in blue with the template overlaid in red.
3.6 Curvature Analysis

In Figure 3.8, although the data itself is extremely messy, we can see from the warping functions $\gamma$ that only slight adjustments must be made by the $\gamma$ function to align these curvature trajectories. Furthermore, the SRSFs in Figure 3.7 seem to be also noisy, with some trajectories having small peaks or deviations in areas other than the major peaks of the original curvature trajectory.

3.6 Curvature Analysis

From these aligned functions, we can compare the ASD and NT participants on a variety of measures. Firstly, we are interested in how they differ in maximum curvature reached. For the first three shapes, the highest curvature occurs at the bottom of the smaller loop. For the ovals, the highest curvature occurs at the top of the oval. Finally, for the figure eights, the highest curvature occurs at the top of the figure eight.
Figure 3.9: Histograms of the maximum curvature in each repetition, split up by shape. We can see a consistent trend of the ASD individuals displaying a higher maximum curvature on average, as well as a wider range of maximum curvatures. The exceptions to this observation are in shapes three and four, in which the distributions seem approximately equal (shape 3) or the effect seems to be reversed (shape 4).
3.6. CURVATURE ANALYSIS

A histogram of the maximum curvature for each shape is shown in Figure 3.9. We see a general trend of the ASD participants having a wider spread than the NT participants, specifically on the higher end of curvatures. This could result from ASD participants drawing more angular shapes, thus displaying a diminished ability to adapt to the changing curvature, as compared to the NT participants. In shapes 3 and 4, however, since the shapes are less circular as seen in Figure 3.4, the maximum curvature of the template is generally higher than that of the other shapes. This may contribute to more similarity between ASD and NT participants, since the shapes themselves are already quite angular and so there would be less of a difference between the two groups.

In addition to the maximum curvature reached, we can look at the variance of the curvature trajectory amplitudes in Figure 3.10. For the first three shapes, we see a slightly more spread out distribution for the ASD subjects than for the neurotypical participants, but the centers of these distributions seem roughly the same. This effect is also observed for shape 5, though not for shapes 4 and 6. In shape 4, the ASD participants seem to have a lower variance on the whole. As in the maximum curvature calculation, perhaps because the ASD subjects do not reach the maximum curvature, their overall variance is lower. By contrast, neurotypical participants that reach or overshoot the maximum curvature will have a larger variance. In shape 6, we observe the opposite effect in that the ASD participants seem to have a higher variance than the neurotypical participants. Because shape 6 is essentially a circle, ASD participants do indeed have a larger variance in their tracings for very simple shapes. We observe the same effect for all three figure eight shapes (7 through 9): again, these shapes in general have lower curvatures, so it would make sense that ASD participants would have higher variance. In addition, as previously mentioned, these figures
Figure 3.10: Histograms of the variance in amplitude in each repetition, split up by shape. For shapes 1 through 3 and 5, the centers of the distributions seem approximately equal, but the range seems wider in the ASD group. Shape 4 seems to have a lower variance for the ASD group, while shapes 5 through 9 seem to have a higher variance on average for the ASD group.
are more complex since one must change curving direction multiple times throughout one tracing of the shape, so perhaps ASD participants have more trouble tracing these shapes.

3.7 Karcher Mean Analysis

We can also compute the Karcher mean, which can be thought of as the curve that minimizes the squared distance from the SRSFs $q_1, \ldots, q_n$ to that curve. The formal definition is below.

**Definition 3 (Karcher mean).** For some set of SRSFs $q_1, \ldots, q_n$ corresponding to the functions $f_1, \ldots, f_n : [0, 1] \to \mathbb{R}$, we define the *Karcher mean* as the curve $q$ that minimizes

$$
\sum_{i=1}^{n} d(q, q_i)^2
$$

where $d(q_1, q_2) = \inf_{\gamma \in \Gamma} \|q_1 - q_2 \circ \gamma\|$.

In our case, we take the norm to be the pointwise distance between two curves: since our curvature trajectories are time-discretized to 200 points in the time interval $[0, 1]$, for two SRSF trajectories $q_1$ and $q_2$, we define

$$
\|q_1 - q_2\| = \frac{200}{\sum_{i=1}^{200} (q_1(t_i) - q_2(t_i))^2}
$$

Thus, to calculate the Karcher mean, we choose the $\gamma$ for each SRSF that minimizes $\|q - q_2 \circ \gamma\|$ and then minimize the entire sum.

Computing and plotting the Karcher mean for each shape yields the graph in Figure 3.11. Although there is much noise for shapes 6 and 7 in particular, we do observe a general trend of “cut-off peaks”, which is observed when the neurotypical participants achieve a higher peak in curvature than the ASD participants. For instance, for the peaks in curvature in
Figure 3.11: Karcher mean trajectories for each of the nine shapes. We can observe the cut-off-peaks phenomenon in shapes 3, 4, and 5 in particular.
shapes 3 and 4, the ASD participants do not quite make it up to the peak curvature that the neurotypical participants seem to display. A physical interpretation of this phenomenon is that when approaching a sharp increase in curvature, neurotypical participants are able to register and adapt to this change relatively quickly, while ASD participants are not as quick to adapt to this change and thereby tend to miss these peaks because they are already adjusting to the lower curvature areas after the peak curvature. Shape 6 is the noisiest, but this is likely due to the fact that the curvature does not change much: the figure itself is very close to a circle, so there are no sharp increases or decreases in curvature to closely examine.

As for the figure eight shapes (7, 8, and 9), I can only conjecture that the noise is due to the complex nature of the curvature itself, as well as the innate difficulty in changing the curving direction. For shapes 8 and 9 have distinctly different loops in the top and bottom of the figure eight, which could be easier for participants to keep a sort of rhythm with. However, we still see some of the “cut-off peaks” phenomenon in these two shapes, though this effect is not as clear as in the other shapes, again likely due to the low values of curvature in the template.

3.8 Speed Analysis

Another aspect of the tracing data is the speed at which the shapes were traced. Although the participants were given instructions to trace at one of four speeds (at their own pace, slow, fast, and intermediate), it was up to the subjects to decide for themselves what each of these meant. Because the data came from a tablet with a constant 146 Hz sampling rate, we calculated the average speed per trial by using a simple formula: the distance between two adjacent sampling points over time taken to draw that distance. We also verified the
CHAPTER 3. THE MOTION OF TRACING IN PEOPLE WITH ASD

As seen in Figure 3.12, the ASD subjects on average traced with slower speed than the neurotypical subjects. This led us to explore the variance of amplitudes in tracing, since one would think that a tracing at a higher speed would impact the variance in various ways. On the one hand, tracing at a high speed could lead to more carelessness and higher variance. On the other hand, high speeds may lead to a more circular shape, which would decrease this variance in amplitudes.

To delve into this phenomenon, for each trial, we plotted the average speed for that specific trial as well as the variance in amplitudes for that trial, shown in Figure 3.13. Along the speed axis, we observe the same distribution as in the histogram: the ASD subjects tended to draw at slower speeds, while the neurotypical subjects drew with a wider range of speeds. However, we can see a very clear difference in the variance: while speed and variance seem roughly uncorrelated for the neurotypical subjects, they are slightly negatively

Figure 3.12: Histogram of speed for the neurotypical group versus the ASD group. We observe, on average, a lower speed for the ASD group than the neurotypical group.

well-known $\frac{2}{3}$ power law for both the neurotypical and ASD group, discussed in Appendix Figure B.1.
Figure 3.13: Scatter plot of variance in amplitudes versus speed across all shapes. The ASD group seems to trace with lower speeds and higher variances than the neurotypical group.

correlated for the ASD subjects. In particular, slower speeds and higher variances were more common in the ASD subjects.

The first three shapes are of particular interest, since this shape (the loop-de-loop) has three peaks in curvature. Reproducing the same graph but only for these shapes in Figure 3.14 yields a clearer view of the differences between the two groups of subjects. The speed range is clearly more constricted, and within the 0 to 40 range, we see that the distribution in variance for ASD subjects is much wider than for the neurotypical subjects. With the exception of a few outliers, the neurotypical subjects have almost no variances above 0.05, with most below 0.01 or so. By contrast, the ASD participants display a fairly right-skewed distribution of variances from 0 to 0.08. Similar graphs for shapes 4 through 6 (ovals) and shapes 7 through 9 (figure eights) can be found in Appendix Figures B.2 and B.3.
Figure 3.14: Scatter plot of variance in amplitudes versus speed across shapes 1 through 3, or the loop-de-loop shapes. The effect that ASD group seems to trace with lower speeds and higher variances than the neurotypical group seems to be even more pronounced in these three shapes.

3.9 Identifying Principal Components

The previous sections have outlined many differences between the two experimental groups, but the reasoning behind these differences still remains abundantly unclear. To help distill these reasons, we conducted functional principal component analysis (FPCA) in order to project the data onto a low-dimensional space. For functions $f_1, \ldots, f_n$ sampled at time points $t_1, T$, we arrange these points in a matrix and compute the sample covariance, denoted by $C$, which is a $(T+1) \times (T+1)$ matrix. Then, we perform the conventional singular value decomposition to obtain principal eigenvalues and eigenvectors [17]. We begin with a few definitions.

Definition 4 ($L^2$ norm). The $L^2$ norm, also called the Euclidean norm, between two vectors $x$ and $y$, both of length $n$, is defined as

$$
\|x - y\|_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}
$$
3.9. IDENTIFYING PRINCIPAL COMPONENTS

Figure 3.15: Percentage of variance explained by each principal component in shapes 3 and 4. About fifty percent of the variance can be explained by the first three principal components.

**Definition 5 (Fisher-Rao metric).** For two SRSFs $v_1, v_2$ and a warping function $\gamma$, we have

$$\langle v_1, v_2 \rangle_\gamma = \int_0^1 \frac{dv_1(t)}{dt} \frac{dv_2(t)}{dt} \frac{1}{\frac{d\gamma(t)}{dt}} dt$$

Interestingly, the $L^2$ norm in SRSF space translates to the Fisher-Rao metric in functional space (see proof in Srivastava et al. 2011 [18]). We found that the first three principal components explained 58% of the variance in the trajectories on average across the nine shapes. Figure 3.15 shows the cumulative percentage of variance explained by the first $k$ principal components, for $k$ ranging from 1 to 20, in shapes 3 and 4 (all nine shapes shown in Appendix Figure B.1). These shapes have the largest peaks in curvature by amplitude and show the “cut-off peaks” phenomenon discussed in the Karcher mean section. We do not see a significant difference in the two curves, hinting that these three principal components may represent the same aspect of the data. For shapes 2, 8, and 9, the graphs look similar to those for shapes 3 and 4, while for all other shapes the graph looks irregular due to the low amplitudes of curvature in these shapes (especially in shapes 1 and 7).
Figure 3.16: Histogram of projections of curvature trajectory on each of the first three principal components for shapes 3 and 4. There seem to be compelling differences in distribution, especially in principal components two and three for both of these shapes.
We can project the data along one of the first three principal components and plot a histogram of the result, shown in Figure 3.16. Although we do not know what these principal components represent in functional space, we can tell that there are clear differences between the two populations for shapes 3 and 4. This is particularly apparent in principal components two and three, both of which have sections where only ASD subjects are represented. This could be due to the larger variance as discussed previously, which could correspond to larger variance when projected to these spaces. Similar graphs for all nine figures are shown in Appendix Figures B.5, B.6, and B.7.

3.10 A Diagnostic Method

An interesting application of the previous results is classification. Studies have shown that early detection can make a huge difference in the quality of life for a child on the autistic spectrum [19]. Because tracing shapes is relatively low effort, cost-effective, and quick, we trained a classifier on a combination of the principal component projections we calculated from the previous section to predict. Due to severely limited data, we used a leave-one-out procedure in which we trained the classifier on nine individuals and used the data from the tenth as the test set, yielding a percentage accuracy. The ten percentages were averaged to obtain an overall accuracy rate.

Many classifiers were used, including but not limited to a support vector machine (SVM), stochastic gradient descent (SGD), and random forests (RF), but the SGD classifier was determined to be the best in terms of accuracy. Due to the drastic differences among the various shapes, we ran each shape separately with various combinations of the first three principal components. As expected, the performance of the classifier varied widely as expected due to the small amount of data as well as the variance among the shapes.
Across shapes, combinations of principal components, and variations in the type of classifier used, we obtained a maximum of 90% accuracy over ten replications on shape 4, using the second and third principal components that were noted as distinguishing between the two groups in the previous section. We obtained a slightly lower accuracy of 83% on the same two principal components in shape 3. As expected, the classifier did very poorly on shapes 1, 8, and 9, due to the noise in the data. However, the findings for shapes 3 and 4 nevertheless indicate that a low-dimensional representation of the data can distill the differences between the two groups.

### 3.11 A Generative Model

#### 3.11.1 Formulating the Model

Based on our results from the previous sections, shapes 3 and 4 seem the most interesting to examine and tease out the mechanism behind the differences displayed by the ASD subjects. We ultimately chose shape 4 due to its simplicity, as well as the existence of just two peaks per revolution, as opposed to four peaks in shape 3.

Figure 3.17 illustrates the first three principal component trajectories for shape 4. We can see that the biggest differences in the principal components seem to occur at the beginning, middle, and end of the trajectory. This happens to coincide with the areas of high curvature in the shape (the bottom and top of the oval). As we saw in the Karcher mean analysis, the “cut-off-peaks” phenomenon also seems to occur around these peaks of curvature. We then hypothesized that the ASD subjects had more trouble quickly increasing from low to high curvature and decreasing back down to a low curvature. In other words, they were slower to respond to this change, thus not able to reach the peak before having to decrease again.
3.11. A GENERATIVE MODEL

Figure 3.17: First three principal component trajectories for shape 4. Most of the difference between the neurotypical and ASD principal component trajectories occurs at the areas of high curvature.

to a low curvature. Thus, perhaps the ASD participants differ in their tracking mechanism, causing them to not reach the peak as quickly.

\[ \epsilon g''(t) + cg'(t) = \lambda(f(t) - g(t)) \]  

(3.1)

To this end, we created a control mechanism (Equation 3.1) for the curvature trajectories. The solution to this equation represents a trajectory that tracks \( f(n) \), the template curvature trajectory, with delay \( \xi \). Our objective, for each curvature trajectory in the data, is to fit parameters \( \epsilon, c, \lambda \) such that the solution \( g(t) \) to this equation is as close as possible to this trajectory in the data.
3.11.2 Fitting the Model

To fit each trajectory in the data, we first made some simplifications to the data itself. Firstly, since one of the two peaks in the trajectory occurs at the beginning and end of the trajectory, we decided to focus only on the middle peak, representing the top of the oval. Thus, we restricted our attention to the middle 100 sampling time points of the original 200. We also defined an objective function to measure how close a solution \( g(t) \) is to the actual data. The objective function is a weighted \( L^2 \) norm (weights will be discussed later).

Using this objective function, we then used MATLAB’s \texttt{fminsearch} function to optimize a fit as close as possible to the data trajectory.

However, we quickly realized that the output of the \texttt{fminsearch} varied widely depending on the starting values of \( \epsilon, c, \lambda \), and thus there were many local minima with respect to the objective function. Thus, for each curvature trajectory in the data, we ran the algorithm with 125 different initial values (5 for each parameter) and chose the fitting with the lowest objective function value.

Furthermore, due to the noise in the data, the fitting of the peak was often poor and computationally expensive. Thus, we performed a Gaussian smoothing of the data with...
a window of 25 time points to eliminate some of the noise in the data. Examples of the smoothing are shown in Figure 3.18.

Interestingly, the parameters \((\epsilon, c, \lambda)\) we obtained from the \texttt{fminsearch} had the fascinating property that the quantity \(\xi\) was constant for each curvature trajectory in the data. In other words, though the actual values of \((\epsilon, c, \lambda)\) may be widely different across initial values, this particular quantity seemed to stay constant for each curvature trajectory.

\[
\epsilon g''(t) + \alpha g'(t) = \beta (f(t) - g(t))
\]

(3.2)

Therefore, to simplify the equation and shrink the parameter space, we rewrote the equation as in Equation 3.2 where the ratio \(\tau = \frac{\beta}{\alpha}\) remains constant. Thus, 125 combinations of initial parameters was reduced to 25, which was much more efficient in terms of computational cost and time. We then calculated this ratio for each trial for shape 4.

For a better visualization of the dynamics of this model, refer to Figure 3.19 for an illustration of how different values of \(c\) can affect the solution to the differential equation. In this particular example, we used a sine curve for \(f(t)\) and fixed a single value of \(c\) and \(\lambda\). As we can see, increasing \(c\) increased the delay in the peak.

We also found that more clarification was needed to quantify how “good” a fitting was. See Figure 3.20 for an example of fitting a curvature trajectory from the data. The same curve (in blue) is fitted, but the two curves in green are different optimizations for fitting. The fitting on the left represents a smaller difference in peak curvatures, but the figure on the right has more aligned peaks. Since we are particularly interested in the delay parameter, we choose to optimize for the aligned peaks so we can quantify how delayed the peak is in comparison with the template. Thus, the objective function gives a higher weight to points on the two slopes of the peak, but not the location of the peak itself.
CHAPTER 3. THE MOTION OF TRACING IN PEOPLE WITH ASD

Figure 3.19: Illustration of how the $c$ parameter affects the dynamics of the model in Equation 3.1 using a sine curve. We see that an increase in $c$ increases the delay in tracking of the curve $f$.

Figure 3.20: Examples of fittings prioritizing different objectives. Left: the solution peak is closer in height to the data, but the peaks of the two curves are not aligned. Right: the solution peak is further away in height to the data, but the peaks of the two curves are aligned. Since our aim is to capture the delay in the curve, we prioritize the fitting of the graph on the right over the graph on the left.
3.11. A GENERATIVE MODEL

Figure 3.21: Boxplot of $\tau$ values for the neurotypical and ASD groups. The ASD group has on average lower $\tau$ values than the neurotypical group, indicating a higher delay in tracking.

Using the simplified equation, we found that ASD subjects had significantly lower values of $\tau = \frac{2}{a} = \frac{1}{c}$, which indicates a longer delay in tracking as compared to the neurotypical subjects. A boxplot of the $\tau$ values is shown in Figure 3.21. In addition to having lower $\tau$ values, the distribution for the ASD group seems to be more narrow, but has more outliers than the NT group. Graphs for specific speed/template settings are shown in Appendix Figure B.8, with further analysis for shape 5 shown in Appendix Figure B.9.

Conducting a two-sample $t$-test for a difference in means yields a $p$-value of below 0.001, which reveals a statistical significance in the difference between the $\tau$ quantity for the ASD and neurotypical participants. The results from fitting our proposed mathematical model suggest that the ASD subjects display more delay in tracking a template curve, particularly in areas of high curvature. However, after reducing an infinite parameter space to a one-dimensional one, it remains a question whether this parameter is enough to capture the wide array of differences we observed in previous sections.
CHAPTER 3. THE MOTION OF TRACING IN PEOPLE WITH ASD

3.11.3 Testing Validity of the Model

To determine the extent to which the parameter $\tau = \frac{\beta}{\alpha}$ captures the difference in tracking delay between the ASD and neurotypical subjects, we generated a simulated dataset and performed the same analysis as in the previous sections to see if we would see the same results. Using our fitted values of $\tau$ from the data, we used the solution from Equation 3.2 and added noise to mimic the actual data. The noise was probabilistically inserted at each time point (Gaussian with small variance). Sample curves with and without noise are shown in Figure 3.22. Then, we divided this dataset into two halves: high values of $\tau$ and low values of $\tau$, regardless of the experimental group they came from.

The two results we wanted to replicate in particular were the Karcher mean \textit{"cut-off-peaks} phenomenon as well as the larger variance in the ASD group (low $\tau$ values, high delay) as compared to the neurotypical group (high $\tau$ values, low delay). However, taking just the best fitting from each individual for each speed and template setting did not yield enough data to observe the phenomenon, since the added noise still seemed to dominate the Karcher mean trajectories. Thus, we took multiple pairs $(\alpha, \beta)$ from each individual for each speed and template setting, perturbing each of the resulting curvature trajectories with random noise to augment the dataset. Then, taking the Karcher mean trajectories
3.11. A GENERATIVE MODEL

Figure 3.23: Karcher mean trajectory of simulated data. The long delay group displays the same cut-off-peaks effect that we saw in the actual data.

across the two simulated groups yielded the same “cut-off-peaks” phenomenon we found in the data, seen in Figure 3.23. Interestingly, the long delay group displayed an even bigger gap at the peaks than the actual data did. This is probably due to the method we used to split the \( \tau \) values into our two simulated groups: there were certainly individuals in the ASD group with high \( \tau \) values in certain trials, as there were also individuals in the neurotypical group with low \( \tau \) values in certain trials. Thus, dividing the pooled \( \tau \) values into its upper and lower halves only made the effect more pronounced.

Figure 3.24 shows the variance in the amplitudes of the simulated data. As we observed in the real data, the long delays group had a higher variance in amplitudes compared with the short delays group. Note that the shapes of these two distributions vary from the ones obtained from the real data, shown in Figure 3.10. This could again be due to the way in which we split the \( \tau \) values. However, unlike the Karcher mean results, the effect seems to be smaller here. This could just be due to random variation, but it could also hint at another mechanism driving this particular observed difference. Specifically, because some of the
Figure 3.24: Variance in amplitudes for simulated data. The long delay group seems to display a higher variance in amplitude, on average, than the short delay group.

neurotypical $\tau$ values were used in the “long delays” group and some of the ASD $\tau$ values were used in the “short delays” group, the two distributions could have been somewhat averaged to produce the simulated variance distribution, shrinking the observable difference between the two groups. However, we still do see the effect of higher variances for the ASD group on average, so perhaps the delay mechanism combines with some other mechanism to produce this difference in variance of amplitudes. From reproducing the differences we observed in the original data, we have demonstrated that the parameter $\tau$, representing the delay in tracking, sufficiently captures those observed differences in curvature trajectories.

3.12 Perception and Behavior

While our analysis so far has been confined to the motor realm, the same group of participants completed various associated cognitive tasks as well. In particular, participants in the study were shown multiple video clip of a dot moving on a screen while tracing simple
3.12. PERCEPTION AND BEHAVIOR

Figure 3.25: Scatter plot of perception task versus variance in amplitudes. We observe a high correlation, indicating the two tasks may be driven by a common mechanism.

shapes, just as they had done in real life. They were then asked if they perceived the dot as a human or a foreign object. The participant’s responses were assessed and a measure was calculated to measure their visual perception of the dot’s motion, as well as an overall variance in response $\sigma_{vis}$. Comparing this measure of variance in visual perception to the variances we calculated above in real motoric action, we calculated the Spearman correlation coefficient to be 0.73 and a scatter plot is shown in Figure 3.25 (a similar plot split by group is shown in Appendix Figure B.12). Using a linear regression $t$-test, we obtained a $p$-value of less than 0.001, indicating that there is compelling evidence of a positive correlation between the two measures. This strong correlation suggests that a common mechanism may drive the two complementary systems of perception and action. Scatter plots showing this relationship appear in Appendix Figures B.10 and B.11.
3.13 Discussion

In this chapter we began by analyzing tracing data from an ASD participant group and a neurotypical participant group. Because the data itself was prone to huge variance, we conducted thorough analysis of the curvature trajectories to distill the data into its key features. We then used functional analysis to align the curvature trajectories, which revealed a “cut-off peaks” phenomenon in the Karcher mean trajectories. This suggested that, compared to the neurotypical participant group, the ASD participants had more trouble tracking and adapting to the sharp increases of curvature, especially in shapes 3 and 4, in which these increases were particularly sudden. In other words, the ASD group appeared to have a larger tracking delay, leading them to not quite reach peak curvature levels before moving on with the rest of the shape. Additionally, the ASD group had a larger variance overall in amplitudes of their curvature trajectories, which indicates that it may be harder for them to track the shape in general. Comparing these quantities to the speed with which the two groups traced the shapes revealed that the ASD group both traced with lower speeds and higher variances than the neurotypical group. Finally, functional principal component analysis (FPCA) revealed that the differences between the two groups could be reasonably compressed into just three principal components. These three principal components were then used as features for a stochastic gradient descent (SGD) algorithm that, for one shape and combination of these principal components, reached 90% accuracy in classifying the test data.

With these findings in mind, we formulated a mathematical model that hoped to capture these key differences, particularly the delay in tracking as seen with the Karcher mean analysis. During the process of fitting this model, we discovered that just one key parameter $\tau$ seemed to govern the behavior of the model, but it still remained a question if this
3.13. DISCUSSION

parameter was responsible for producing the delay in tracking we had observed in our earlier analysis as well as other differences we observed and that we were attempting to replicate in the model. To that effect, we created a simulated dataset using only this parameter and split this dataset into high and low values of $\tau$, corresponding to short and long delays, respectively. Replicating our curvature analysis that we performed on the original data with our new simulated dataset produced the same differences we observed in the data, indicating that this delay parameter $\tau$ did seem to be sufficient in reproducing these differences, and in some cases the simulated dataset produced a larger difference, especially in the Karcher mean “cut-off peaks” phenomenon. Finally, we saw evidence that there may be a common mechanism that engenders the differences between the ASD population and neurotypical individuals in both perception of motion and actual motion.
Chapter 4

Discussion and Conclusion

4.1 Connecting Two Forms of Behavior

In the previous two chapters, we have seen two projects that differ greatly in both their scale and purpose. In the former, we observe a small organism’s efforts at exploring and interacting with a large environment, motivated by the need for resources, while in the latter, we witness a task that is not at all necessary for living but perhaps has larger implications. We identify different quantities that allow for an interesting lens through which we can examine our motion data: the fly’s distance from the obstacle in relation to various aspects of the fly’s path in the former and the curvature trajectory of the tracing data that allowed us to distill the noisy tracing data in the latter. However, both research questions are similar: how can we take our numerical analysis of the geometries and various features of these two types of motions and use it to devise a generative model that incorporates and hopefully reproduces these traits that we highlighted in our initial analysis? Moreover, given this
proposed generative model, how do we evaluate its quality, and how does its performance inform our understanding of the dynamics and geometry of the motion?

In the *Drosophila* project, from our initial analysis, we hypothesized that there were two modes of movement that governed the fly's behavior in this two-dimensional walking environment: an exploring behavior characterized by large changes in heading and slower speeds as well as a traveling behavior characterized by mostly straight walking and higher speeds. We then formulated a mathematical model that incorporated these findings and tested it by both generating a simulated data set and performing the same analysis to see if we obtained similar results. The same methodology was applied to the ASD project, in which we identified an apparent delay in tracking for the ASD population and built a mathematical model that incorporated tracking delay as a parameter. In the same fashion as the *Drosophila* project, we then generated a simulated data set and reproduced the results that we found were most important in differentiating between the ASD and neurotypical populations.

In both projects, our aim was to identify the key features of the particular movements we investigated, but how would we know that our proposed key features were indeed what produced interesting characteristics in the data analysis? We evaluated the caliber of our model by generating simulated data and demonstrated that this simulated data still had these characteristics, indicating that it was likely the features that we pinpointed, not other aspects of the data, that were responsible for producing these characteristics. Moreover, the success of our models provides evidence that these key features play a large role in governing the dynamics of not only these two movements but also possibly for other documented behaviors. Specifically with the ASD group, we found a significant positive correlation between their cognitive perception of movement and their movement itself.
4.2 Limitations

However, the fact that these models reproduced the desired characteristics does not necessarily establish a causal or direct relationship between the identified key features of the data and these characteristics, but merely provides some evidence, albeit quite strong, that there is some nature of a relationship between the two.

In the Drosophila project, the data itself was extremely noisy and each trial had its own flavor in terms of how likely the fly was to travel or stay in place, as well as how much of the arena the fly explored. Moreover, finding segments of the paths across trials in which the fly only interacted with one object was not possible, since we cannot know whether the fly noticed other objects in the vicinity and chose to not explore them for various reasons or if the fly simply did not see the other objects. Thus, the size of the dataset for this analysis was small, meaning that the characteristics we observed could just be due to chance or other factors that caused the data to take on these characteristics. However, due to the different natures of each trial and across flies, the fact that we did observe the general trend of having two modes of movement, as well as our analysis of how distance from an obstacle impacts movement even with these drastically different trials indicates that this general trend does seem to appear in fly movement.

For the ASD project, perhaps the biggest limitation in our analysis is the small data set. Although the number of individual repetitions for each participant was quite large (on the order of $10^5$), the number of participants in each group was still limited to just ten. As much of the literature on ASD has shown, the manifestation of autism in each individual varies quite widely, so there is no guarantee that these ten people are representative of the ASD population as a whole. Along the same lines, the various differences observed in the distribution of variances, speeds, values of $\tau$, etc., were relatively small, though most were
statistically significant. Again, because of the existence of a spectrum for both autism as a disorder as well as the appearance and severity of symptoms, much of the distributions overlapped and differences were relatively small. However, part of the challenge of this project was to tease out what factors could be behind these small differences, and we did find that the simulated data set, split up based on high and low $\tau$ values, produced a larger effect than the data itself.

4.3 Directions for Future Work

While we have both identified and shown the significance of the various features of interest we identified in each scenario, many questions remain about our understanding of these features and the overall geometry of the movement itself, as well as how the generative models themselves can be improved.

4.3.1 Flies

Much remains to be explored in the *Drosophila* project: first and foremost, the model could be extended to a multi-obstacle environment. Previous research showed that a fly’s decision to move towards an object was relatively unaffected by the surrounding objects in the environment, but questions persist as to how the fly chooses the obstacle in the first place [3]. We could incorporate these findings into our mathematical model that could reveal other aspects of the fly’s movement that have larger implications. Along these lines, we could delve into a fly’s memory of obstacles that it has or has not explored. Specifically, we could ask questions about how a fly would choose its next obstacle between one that has been explored but perhaps had resources near it and one that is as of yet unexplored. Additionally, many of the flies traveled to the edges of the arena, past the grid of obstacles.
4.3. DIRECTIONS FOR FUTURE WORK

This is another form of obstacle that can be studied: what happens when a fly comes across a flat surface and must turn back? Will the fly return to that obstacle it came from, or will it set off in a completely new direction?

Furthermore, some of the data had the interesting quality that the flies were walking in a visually empty arena, though the obstacles were still physically present. We could look at differences in behavior between flies walking in each of these two setups; we would expect that the flies would not exhibit as much of a decrease in speed if they cannot see the obstacle. In fact, perhaps the radius of distances at which the fly slows down would be much smaller, since the fly would only explore the object after bumping into it. After exploring this seemingly invisible object, we could imagine that the fly would then change its modes of movement to be perhaps more cautious or slower while walking, but it could also continue exploring other areas as before.

Another dimension of interest would be the presence or absence of optogenetic triggers around the obstacles that either attract or push away flies from that particular obstacle. This adds a dimension of attractiveness or repulsiveness of obstacles that would affect our model of how flies interact with these obstacles, as well as how flies decide to explore particular objects, specifically when it has associated a particular visited obstacle with either attractive or repulsive qualities and has perhaps stored this information in its memory as it explores.

Based on the setups of these trials, we could then perform analysis similar to our work in the ASD project in identifying key features that could relate to any differential characteristics that we observe in one scheme but not the other. For instance, the sharpness of the transition between the exploring and traveling modes of movement could differ in a visible or invisible obstacle setting. This analysis would be greatly benefited by the ability to fit parameters for the data. Currently, the parameter space for our model is very large and finding a set of
best-fit parameters is computationally expensive. As with the ASD project, we can attempt to reduce this parameter space to just one or two parameters, which would offer a great deal of benefits. Moreover, we could fit parameters for each trial in the data that would better quantify how the trials differ from one another as well as how each group (visible versus invisible obstacles, various optogenetic stimuli) could fundamentally differ from one another. This could reveal the cornerstones of the dynamics of fly movement, as well as how flies process information and integrate it into their movements.

4.3.2 ASD

Our mathematical model was only trained on shape 4, as it was deemed one of the most differentiating shapes for tracking delay between the ASD and neurotypical groups. However, as we have observed in many of our analyses, the performance of these two groups varies quite widely based on shape and speed settings, as seen in Appendix Figures B.8 and B.9. Opposite effects were even observed in some analyses. Thus, a natural extension of the current work would be to examine the other shapes, starting with the other ovals (shapes 5 and 6) and perhaps expanding into the loop-de-loops (shapes 1 through 3). The latter shapes are quite similar to the ovals and have similar curvature trajectories, though they differ in that the participant must draw loops of alternating sizes, which provides yet another layer of complexity to the task. The figure eights (shapes 7 through 9), however, may require an entirely different model due to their unique curvature trajectories and the need to change directions throughout each tracing of the shape.

Furthermore, analysis of various speed/template settings within each shape could be illuminating: we would expect that the ASD population performs better at slower speeds due to their demonstrated attention to detail and therefore worse at higher speeds. Also, the
4.3. DIRECTIONS FOR FUTURE WORK

ASD population could perform worse without a template due to a demonstrated difference in the ability to incorporate past information with current tasks. However, much more data is needed for this particular analysis, as the number of each individual’s repetitions within each speed/template setting is relatively small and results could perhaps be due to noise and thus unrepresentative of the population.

Moreover, our analysis of the principal components revealed that the data could be significantly compressed into just three principal components. Much still stands to be learned about what specifically these principal components are and what they manifest as in the tracing space. If we can reconstruct the shape drawn from each of the three principal components per shape, we can compare the results from both groups to identify where variance in each group is coming from. This would also better inform our comparison of the performance of both groups in areas of high or low curvature, as well as inform the parameters and form of our mathematical model. Knowing the specific aspects of the shape that provide differences between the two groups would also be helpful as inputs to the classifier we trained on the principal components.

An interesting spinoff of this project is the classifier that we trained with the first three principal components from FPCA. Though not much time was spent on this, it would be useful if the classifier were more robust and could be trained on more data. Moreover, only FPCA results were used for this classifier, but other components could be added, such as various parameters from our model or other features from numerical analysis of the data. A successful classifier could be hugely useful for early detection of ASD, as drawing variants of circles is not a particularly complex task and is reasonably fast and cost-efficient. However, a much larger quantity of data would be needed for a tool that could be deployed to the youth population of the world. Moreover, as more findings emerge about distinctive
characteristics of individuals on the autistic spectrum, more features could potentially be used in the classification. As mentioned previously, diagnostic tools for ASD widely vary due to the diverse nature of this disorder. This may necessitate a regression in place of a classifier that predicts a number on a scale of 0 to 1 rather than a binary result. A range of possible results would mesh well with current measures, such as the Autism Quotient, that also provide submeasures that capture the wide-reaching implications of ASD.

Finally, the mechanisms driving motor behavior could potentially be linked to mechanisms driving perception in the ASD population. While we have shown that two specific behaviors could be linked, much remains to be uncovered about the particular structures behind our observations of differences in both mental and physical behavior. Linking the underlying mechanisms of behavioral differences with those of perception could provide a stronger underpinning for our understanding of the disorder as a whole in its widely varied manifestations in the ASD population and could be instrumental to better diagnosis and treatment.

To conclude, we have characterized the dynamics of two forms of movement: *Drosophila* walking in a two-dimensional VR environment and people on the autistic spectrum tracing shapes on a tablet. In the *Drosophila* project, we identified two modes of movement, a local search mechanism as well as an exploitative traveling behavior, as likely characterizing a fly’s movement in a grid of obstacles. These two movements displayed different characteristics in the speed, angular velocity, and curvature of the movement. A simulated data set from the generative model displayed the same trends we observed in the original data, suggesting that our identified features play a key role in characterizing this behavior.

Similarly, in the ASD project, we identified tracking delay as a key parameter that served as a basis for the differentiating factors observed in various aspects of the tracing
4.3. DIRECTIONS FOR FUTURE WORK

Through the generative model, we demonstrated that differentiating based on the degree of tracking delay produced similar differences in our numerical analysis as the ASD and neurotypical groups did, suggesting that tracking delay plays a role in the fundamental difference in performance of the two groups in this task.

Though these two forms of movement exceedingly differ in their physical scale and mechanism, we have shown that their geometries can be analyzed and characterized through the formulation of a generative model that both informs and strengthens our understanding of the movements. This process facilitates our comprehension of the role various factors play in driving these movements and provides us with a method to quantify the strength of these effects. Our analysis could provide a basis for understanding other facets of these organisms, such as modes of thinking and perceiving. There remain innumerable facets of movement in this world that are still waiting to be understood, as well as many more that are as of yet undiscovered. This method of analyzing and characterizing the movement through generative models is merely one tool, but its wide applicability can allow us to better discern the subtleties of movement that can illuminate and guide our understanding of this world.
Bibliography


Appendix A

Additional Drosophila graphs

This section contains additional graphs for the Drosophila project.
Figure A.1: Additional simulated paths according to Equation 2.1. For two of the paths, the fly leaves but turns around and comes back to the obstacle. All have the property that speeds near the obstacle tend to be slower than speeds farther away from the obstacle. Furthermore, the angular change closer to the obstacle seems to be larger on the whole than the angular change further away from the obstacle.
Figure A.2: Histograms of the ratio of the deterministic term to the error term for position and heading, split up based on short or long distances away from the obstacle. For all three of $x$, $y$, and $\theta$, the ratio is larger for long distances and smaller for short distances, implying that the error term is smaller. This is what we expect, since the deterministic term should dominate at further distances, while the noise term should dominate at short distances.
Appendix B

Additional ASD graphs

This section contains additional graphs for the ASD project.
Figure B.1: Demonstration of the $\frac{2}{3}$ power law for all nine shapes. This law says that for a human drawing a figure, the relationship between the logarithm of the curvature and the logarithm of the speed are related linearly with a slope of $-\frac{1}{3}$. This result is shown for both the neurotypical and ASD groups in the figure above.
Figure B.2: Scatter plot of variance in amplitudes versus speed across shapes 4 through 6, or the oval shapes. The effect that the ASD group seems to trace with lower speeds and higher variances than the neurotypical group seems to be even more pronounced in these three shapes.
Figure B.3: Scatter plot of variance in amplitudes versus speed across shapes 7 through 9, or the figure eights. The effect that the ASD group seems to trace with lower speeds and higher variances than the neurotypical group seems to be not as pronounced and almost nonexistent in these three shapes.
Figure B.4: Percentage of variance explained by each principal component in all shapes. About fifty percent of the variance can be explained by the first three principal components. In shapes 1, 6, 7, and 8, the ASD curve seems to deviate significantly from the neurotypical curve, but this could be due to the fact that the data in these four shapes is particularly noisy and has relatively low curvature in general.
Figure B.5: Histogram of projections of curvature trajectory on each of the first three principal components for the first three shapes (the loop-de-loops). There seem to be compelling differences in the distribution for all three principal components across all three shapes.
Figure B.6: Histogram of projections of curvature trajectory on each of the first three principal components for shapes 4 through 6 (the ovals). There seem to be compelling differences in the distribution for the second principal component across all three shapes, and the third principal component in shape 6.
Figure B.7: Histogram of projections of curvature trajectory on each of the first three principal components for shapes 7 through 9 (the figure eights). There seem to be compelling differences in the distribution for the third principal component across all three shapes, and the second principal component in shapes 7 and 9.
Figure B.8: τ values for shape 4, based on four separate speeds. Speed 3 is with the template at a slow speed, speed 4 is without the template at a slow speed, speed 5 is with the template at a fast speed, and speed 6 is without the template at a fast speed. As we can see, the biggest differences between the two groups is observed in speed/template settings 3 and 6.
Figure B.9: $\tau$ values for shape 5, based on four separate speeds. Speed 3 is with the template at a slow speed, speed 4 is without the template at a slow speed, speed 5 is with the template at a fast speed, and speed 6 is without the template at a fast speed. There is a much smaller difference, almost negligible, between the two groups. This could be due to the smaller changes in curvature, which is what we suspected would produce the differences in the ASD group.
Figure B.10: Scatter plot of the block design task versus variance in amplitudes. In the block design task, the participants move blocks of colors around to match a specific pattern. The scatter plot shows a negative relationship between the two and a relatively low $p$-value for this correlation.
Figure B.11: Scatter plot of the matrix reasoning task versus variance in amplitudes. The matrix reasoning task asks participants to observe a grid of shapes, identify the pattern, and fill in the empty slot in the grid. The scatter plot shows a negative relationship between the two and a relatively low $p$-value for this correlation.
Figure B.12: Scatter plot of the motion perception task versus variance in amplitudes, but split up by group. We see a positive relationship between the two, but the strength of the correlation is much higher for the ASD group than the neurotypical group.