The visual perception of 3D shape from stereo: Metric structure or regularization constraints?

Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of the Ohio State University

By
Ying Yu, B.A.
Graduate Program in Psychology

The Ohio State University

2017

Thesis Committee:
Alexander A. Petrov, Advisor
James T. Todd
Brandon M. Turner
© Copyright by

Ying Yu

2017
Abstract

A substantial number of psychophysical studies have found that the visual perception of 3D shape from stereo is systematically distorted so that the perceived shape is progressively compressed along the depth dimension as the viewing distance increases. This suggests that 3D shape perception depends on the perception of 3D metric structure, which requires the depth magnitude information derived from stereo to reconstruct the local structures, such as angles and line segments. These studies, however, have been criticized by Pizlo (Pizlo, 2008; Pizlo, Li, Sawada, & Steinman, 2014) for using impoverished stimuli that are not sufficiently well-structured to allow the application of powerful regularization constraints such as symmetry, planarity and compactness. His model, which takes these constraints as prior cues and recovers 3D shapes in a holistic way, can reliably reconstruct the shape from a single projection image. In Pizlo’s model, the depth order information derived from stereo is the only binocular cue needed to improve the reconstructed shape to a nearly veridical level. The above two theoretical positions differ in many respects and we want to know which one can better account for the mechanism underlying people’s 3D shape perception from stereo. Two experiments were conducted to evaluate people’s ability to disambiguate 3D shapes defined by binocular disparity within the ambiguity family that formed by stretching a 3D shape in depth (i.e., Z-scale family) using different stimuli and different tasks under different viewing conditions. Although observers
were able to make reliable judgments, most of them revealed systematic failures of shape constancy over changes in viewing angles and viewing distances. These findings can be largely accounted for by the misperception of metric structure and cannot be fitted by Pizlo’s model. A few observers in Experiment 2, however, produced a different judgment pattern that cannot be explained by the misperception of metric structure and is somewhat aligned with the prediction of Pizlo’s theory. The observed dissociation suggests the coexistence of two strategies in 3D shape perception: an analytical strategy based on local metric structure and a holistic strategy based on global configuration.
To my beloved parents & husband
Acknowledgments

After six months of continuous effort and struggle, this thesis has finally touched to its end. This painful but joyful journey has intensively boosted the strength and growth of me, not only in the scientific arena, but also on a personal level. I would like to express my sincere gratitude to those who have been with me throughout this journey. Without them, this thesis would have never been possible.

I would first like to thank my advisor, Dr. Alexander Petrov. Alex, you are my academic father who leads me into the academic world with patience, caring, respect, discipline and solid instructions. I have learned so much from those inspiring and lengthy conversations between you and me over the past three years. I have been immensely motivated by your enthusiastic craving for truth and abiding passion for knowledge. To be an independent researcher like you is the goal that I am striving for.

I would also like to thank other professors in the Department of Psychology, such as Dr. Jay Myung, Dr. Trisha Van Zandt, Dr. James Todd, Dr. Brandon Turner, Dr. Laura Wagner, Dr. Per Sederberg, Dr. Julie Golomb, and etc., who help to broaden my knowledge in mathematical analysis, vision, cognitive models and interdisciplinary methodology. All the knowledge I learn from you have equipped me to a scientific researcher in the field of psychology.

Special thanks go to my committee members, Dr. James Todd and Dr. Brandon Turner, for their extra contributions to the present project. Jim, your excellence
guidance is an indispensable force pushing this research forward in the right direction. I am also grateful to your generous offer of the experimental equipments and lab space. Additionally, I appreciate your caring, encouragements and constructive suggestions during those dark days when I was struggling with the thesis writing. You did such a great job in helping me out. Brandon, thanks for your help and insight into the hierarchical Bayesian analysis, which is the major statistical tool for the current research. I am also touched a lot by your generous encouragements and kind understanding along the way.

I owe a debt of gratitude to Dr. Zygmunt Pizlo, a professor in the Department of Psychological Sciences at Purdue University, for his highly professional manner in dealing with scientific criticisms. Although there exist some disagreements on how visual system works between you and us, you are always so supportive and willing to exchange ideas with us. I also need to acknowledge Dr. Tadamasa Sawada, a collaborator of Dr. Pizlo, for generously providing us with the original c++ code of making the truck-like stimuli on which the stimuli of our Experiment 2 were built.

I am grateful for all the participants of the current research. It is your kind help that makes this project happen.

Finally, I would like to say “Thank you and I love you” to my parents and husband. You have given me unconditional love and endless supports throughout the years. You trust me much more than I trust myself and have never ever given up on me. You have saved me from those bleakest moments and made me strong. Without you, I would have never been me. Now, I am on my way of being a better person because of you, and for you.
Vita

2010 ................................. Tianjin No.1 High School

2014 ................................. B.A. Psychology, Nankai University

2014-Present ........................ Graduate Teaching Associate, Department of Psychology, The Ohio State University

Publications


Fields of Study

Major Field: Psychology
# Table of Contents

Abstract ......................................................... ii
Dedication ......................................................... iii
Acknowledgments ................................................... v
Vita ................................................................. vii
List of Tables ...................................................... x
List of Figures ...................................................... xi

CHAPTER PAGE

1 Introduction ..................................................... 1
   1.1 Experimental Data on the Systematic Distortion of 3D Shape Perception from Stereo ........................................ 2
   1.2 Pizlo’s Criticisms and His Model .................................. 6
   1.3 Experimental Ideas ............................................. 12
   1.4 Bayesian Inference ............................................. 15

2 Experiment 1: Perceptions of the Shape of Squares ............. 17
   2.1 Methods ...................................................... 18
       2.1.1 Participants ............................................ 18
       2.1.2 Stimuli .................................................. 18
       2.1.3 Apparatus .............................................. 21
       2.1.4 Procedure ............................................. 21
       2.1.5 Analyses ............................................... 24
           2.1.5.1 Bayesian ANOVA analysis ......................... 24
           2.1.5.2 Hierarchical Bayesian analysis .................. 26
   2.2 Results ..................................................... 30
       2.2.1 Preliminary Results .................................. 31
       2.2.2 Main Results ......................................... 33
           2.2.2.1 Ocularity Effects ............................. 35
2.2.2.2 Viewing Angle Effects .......................... 38
2.2.2.3 Consistency and Stimulus Size Effect ............ 38
2.3 Conclusions and Discussion .......................... 40
3 Experiment 2: Perceptions of the Symmetry of 3D Polyhedra .... 46
   3.1 Methods ........................................... 49
      3.1.1 Participants ................................... 49
      3.1.2 Stimuli ....................................... 49
      3.1.3 Apparatus ................................... 53
      3.1.4 Procedure ................................... 54
      3.1.5 Analyses ..................................... 55
         3.1.5.1 Bayesian ANOVA analysis ................. 56
         3.1.5.2 Hierarchical Bayesian analysis .......... 56
   3.2 Results ........................................... 59
      3.2.1 Preliminary Results ........................... 60
      3.2.2 Main Results ................................ 60
   3.3 Control Experiment: Partial Shape .................. 65
   3.4 Conclusions and Discussion ........................ 72
4 General Discussion ....................................... 80
References ............................................... 84
Appendix A: The Stimuli of Experiment 2 ................... 95
List of Tables

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>32</td>
</tr>
<tr>
<td>3.1</td>
<td>59</td>
</tr>
<tr>
<td>3.2</td>
<td>67</td>
</tr>
<tr>
<td>3.3</td>
<td>69</td>
</tr>
</tbody>
</table>

2.1 The top five models for fitting the data of Experiment 1 selected by the Bayesian ANOVA analysis according to the descending order of the Bayes factor values. The model with the greatest $BF_{10}$, i.e., the model that ranked the first, is the best.

3.1 The four candidate models in the Bayesian ANOVA analysis of the main study of Experiment 2. They are ranked from the best (#1) to the worst (#4) according to the the descending order of the Bayes factor values.

3.2 The four candidate models in the Bayesian ANOVA analysis of the control study of Experiment 2. They are ranked from the best (#1) to the worst (#4) according to the the descending order of the Bayes factor values.

3.3 The top four candidate models in the Bayesian ANOVA analysis of Experiment 2, combined the main study and the control study. Only the eleven participants who completed both studies were included in this analysis. The models listed here are the top four in fitting the data. The model that has the largest $BF_{10}$ (model #1) is the best one.
List of Figures

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>An illustration of the stimuli in Experiment 1. The stimuli were computer-generated parallelograms displayed on a gray background. Each of the displayed parallelograms was generated from the following two steps in sequence. First, rotate a basic square around its center within the screen plane from the canonical position (namely, position of the first parallelogram of each rows) by a random degree. This illustration chose 45° and 90° as examples. Then stretch or compress this rotated square along the horizontal axis of the displayed screen (i.e., X-axis) by a factor randomly chosen from the range of (0.5, 2). We call this factor as width-to-height ratio (i.e., W2H), which is defined by the ratio of the displayed parallelogram’s extent along the X-axis to the parallelogram’s extent along the Y-axis (namely, the vertical direction of the displayed screen). The first row applied W2H=1 to the basic square, leading to squares with no shape distortion across different rotated positions. Whereas the second row applied W2H=0.67 to the basic square, leading to different shapes at different rotated positions. It thus can been seen that the two-step procedure of making the stimuli ensures that the rotation of a displayed non-squared parallelogram (W2H≠1) around its center forms a non-rigid motion. . . . . . . . . . . . 20</td>
</tr>
<tr>
<td>2.2</td>
<td>An illustration of the experimental layout in Experiment 1. . . . . . 22</td>
</tr>
<tr>
<td>2.3</td>
<td>Three levels of viewing angle: zero, half and full, were used in Experiment 1. The experimenter manipulated the viewing angle before each block by manually slanting the upright monitor in depth about its vertical midline by 0° (zero), 38.6° (half) or 67° (full) without tilt. 23</td>
</tr>
</tbody>
</table>
2.4 Graphical model and model specifications for the hierarchical Bayesian analysis for Experiment 1. $y_{ijk}$ referred to the natural log of the observation, adjusted $W2H$, in each trial per subject. $x_{ijk}$ indicates the stimulus size for each trial. 3, 5 and 7 correspond to small, medium and large size, separately. Since the current study has 6 experimental conditions, i.e., 2 (ocularity: binocular or monocular) $\times$ 3 (viewing angles: 0°, 38.6° or 67°), $k$ denotes which condition the data, $y_{ijk}$, belong to.

2.5 Posterior means of the estimated $\log(W2H)$ as a function of the viewing angle for binocular (the red line) and monocular (the blue line) viewing conditions in Experiment 1. The six point values correspond to the posterior means of the $\mu^\theta_k$ ($k=1,2,3,4,5,6$), the parameters of the hierarchical Bayesian analysis indicating averaged participant’s performance in each experimental condition. $\log(W2H)=0$ (denoted by the dashed line in this plot) means adjustments that produce perfect squares. $\log(W2H)>0$ means stretching a square along the horizontal axis of the screen and $\log(W2H)<0$ means compressing a square along this direction. Error bars depicted $\pm$1 standard deviation of the posterior distribution, which is comparable to the standard error of the mean.

2.6 Effect of ocularity on the mean adjustments (measured by $\log(W2H)$) for each viewing angle in Experiment 1. The graphs plot estimated posterior probability distributions for the difference in mean adjustments between binocular and monocular viewing conditions, $\mu^\theta_{mono} - \mu^\theta_{bino}$, for each viewing angle. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution.

2.7 Effect of viewing angle on the mean adjustments (measured by $\log(W2H)$) for binocular and monocular viewing conditions in Experiment 1. The graphs plot estimated posterior probability distributions for the difference in mean adjustments between full angle (67°) case and zero angle case, $\mu^\theta_{full} - \mu^\theta_{zero}$, for each ocularity viewing condition. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution.

2.8 Consistency in performance across different observers in Experiment 1. Ocularity effect, viewing angle effect as well as interaction effect estimated for each observer. Each colored curve represents a posterior probability distribution of the corresponding contrast for each observer. The correspondence between the colors of curves and observers’ identification numbers is shown in the legend.
3.1 An illustration of the polyhedra stimuli used in Experiment 2. Each polyhedron is subjected to the same constraints of the stimuli used in Y. Li, Sawada, Shi, Kwon, and Pizlo (2011). It has one and only one invisible symmetry plane and it is mirror symmetrical with respect to this plane. The dashed line depicts the intersection of the polyhedron and its invisible symmetry plane. The polyhedron is composed of three convex hexahedra which can be seen as "top", "neck", and "bottom". The back faces, which are invisible in this figure, are coplanar and orthogonal to the symmetry plane. The front faces are the ones that corresponds to the back faces and are not coplanar. The top face refers to the "roof" of the "top" part. And the side faces refer to the ones that have no intersection with the symmetry plane. The top face, all the three front faces as well as the three side faces on one side are visible in this illustration as well as in all the polyhedra used in the experiment.

3.2 Sample stimuli from the main study of Experiment 2. The graphs show two of the polyhedra used in the main study of Experiment 2. Each polyhedron is displayed at three slants (30°, 45° and 60°, from left to right respectively). These are just monocular images, but binocular images (stereoscopic images for the subject’s left and right eyes) were displayed during the experiment.

3.3 Graphical model and model specifications for the hierarchical Bayesian analysis for Experiment 2. $y_{ijk}$ referred to the natural log of the observation, adjusted $S$, in each trial per subject. $k$ indicates the two viewing distances, 103.5 cm and 198.8 cm.

3.4 Estimated log($S$) adjustments with respect to near and far viewing distances in the main study of Experiment 2. Panel a shows the posterior means of the parameters, $\mu^\theta_{\text{near}}$ and $\mu^\theta_{\text{far}}$, used in the hierarchical Bayesian analysis of the main study. These parameters estimate the values of log($S$) adjustments at the group level. Error bars depict ±1 standard deviation of the corresponding posterior distribution, which is comparable to the standard error of the mean. Panel b shows the posterior probability distributions of the above two parameters. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution. The dashed lines denote the ground truth (namely, log($S$)=0).
3.5 Posterior probability distributions of estimated the log($S$) adjustments in the near (curves filled in dark gray) and far (curves filled in red) viewing distances for each observer in the main study of Experiment 2. Each curve represents the posterior probability distributions of a $\theta_{ik}$, a parameter that estimates the log($S$) adjustments for the $i$th observer in the $k$th (either near or far) viewing distance. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution. The dashed lines denote the ground truth (namely, log($S$)=0).

3.6 Sample stimuli from the control study in Experiment 2. The same set of 15 polyhedra as the ones used in the main experiment was generated but only the three front faces of each polyhedron in this set was displayed during the control experiment. The stimuli shown in this figure were derived from the same polyhedra as shown in Figure 3.2. These are just monocular images, but binocular images (stereoscopic images for the subject’s left and right eyes) were displayed during the experiment.

3.7 Estimated log($S$) adjustments with respect to near and far viewing distances in the control study of Experiment 2. Panel a shows the posterior means of the parameters, $\mu_{near}^\theta$ and $\mu_{far}^\theta$, used in the hierarchical Bayesian analysis of the control study. These parameters estimate the values of log($S$) adjustments at the group level. Error bars depicted $\pm 1$ standard deviation of the corresponding posterior distribution, which is comparable to the standard error of the mean. Panel b shows the posterior probability distributions of the above two parameters. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution. The dashed lines denote the ground truth (namely, log($S$)=0).

3.8 Posterior probability distributions of estimated the log($S$) adjustments in the near (curves filled in dark gray) and far (curves filled in red) viewing distances for each observer in the control study of Experiment 2. Each curve represents the posterior probability distributions of a $\theta_{ik}$, a parameter that estimates the log($S$) adjustments for the $i$th observer in the $k$th (either near or far) viewing distance. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution. The dashed lines denote the ground truth (namely, log($S$)=0).
3.9 Posterior means of log(S) adjustments for each observer in the main (figure a) and control (figure b) studies of Experiment 2. Each data point in the graphs represents the estimated log(S) adjustments for one observer, with the adjustments in the near viewing distance corresponding to its x-coordinate and the adjustments in the far viewing distance corresponding to its y-coordinate. The red triangle denotes the ground truth (namely, log(S)=0). The area above the diagonal of each scatterplot is shaded to illustrate the positive effect of viewing distance. Namely, the data point that falls into this shaded area tells the information that its corresponding participant tends to stretch the 3D shapes (either complete or partial) more along the depth than he or she did in the near viewing distance.

A.1 The entire set of stimuli used in the main study of Experiment 2. These are just monocular images, but binocular images (stereoscopic images for the subject’s left and right eyes) were displayed during the experiment.
CHAPTER 1

Introduction

What appears to be a most common and obvious experience of human observers is that people perceive the world as a three-dimensional (3D) scene populated by physical objects with 3D shapes that do not appear to change as people walk around them. From a theoretical point of view, however, the above obvious experience becomes most perplexing since it amounts to a mathematically ill posed inverse problem (Pizlo, 2001; Poggio, Torre, & Koch, 1985). The actual optical inputs for 3D perceptions are 2D retinal projections of the original physical space with massive information loss, especially the information along the depth dimension. The 2D retinal inputs, therefore, are consistent with infinitely many 3D interpretations. In order to obtain a unique interpretation, the brain needs to impose additional information on the 2D retinal inputs and reconstruct the 3D shape. There are two theoretically significant questions related to this process. First, what is the additional information used by the human brain to help the reconstruction of 3D shapes? And second, does the reconstructed percept have the same Euclidean geometry as that of the distal object? In other words, can the human brain perceive 3D shapes veridically?
1.1 Experimental Data on the Systematic Distortion of 3D Shape Perception from Stereo

To answer the above questions, various depth cues have been explored and have conventionally been considered as building blocks of 3D shape perception. They include ocular-motor cues, such as binocular vergence, accommodation and gradient of retinal blur (e.g., Watt, Akeley, Ernst, & Banks, 2005; Foley, 1980), single image cues, such as texture gradients (e.g., Todd, Thaler, Dijkstra, Koenderink, & Kappers, 2007), light field (e.g., Koenderink & van Doorn, 2004; Koenderink, Pont, van Doorn, Kappers, & Todd, 2007), occlusions (e.g., Norman, Todd, & Orban, 2004), familiar size as well as linear perspective (e.g., Gardner, Austerweil, & Palmer, 2010), and multiple image cues, including motion (e.g., Todd & Bressan, 1990; Norman & Todd, 1993) and binocular disparity (e.g., Johnston, 1991).

Among these depth cues, binocular disparity, which refers to the small positional differences between two eyes’ retinal images, probably is the most powerful one. Binocular disparities alone can give rise to a vivid and compelling impression of 3D structure (Julesz, 1971). More importantly, binocular disparity has been found to be more effective than other depth cues, such as motion (Frisby, Buckley, & Duke, 1996; Durgin, Proffitt, Olson, & Reinke, 1995; Rogers & Collett, 1989; Tittle & Braunstein, 1993), texture (Johnston, Cumming, & Parker, 1993; Tittle, Norman, Perotti, & Phillips, 1998) and shading (Tittle et al., 1998), especially in perceiving relative depths.

A key reason for the effectiveness of binocular disparity is that it can provide both qualitative and quantitative cues to depth. Suppose $A$ and $B$ are two points close to the line of sight. Then the geometric relation among the depth interval ($d$) between $A$ and $B$, binocular disparity ($\Gamma$, exclusively refers to horizontal disparity), fixation
distance \( D \), also known as viewing distance or egocentric distance) and interpupillary
distance \( I \) can be approximated by the following equation (Kaufman, 1974; Foley,
1980; Mon-Williams, Tresilian, & Roberts, 2000; Howard & Rogers, 2002):

\[
d \approx \frac{\Gamma D^2}{I}
\]  

(1.1)

when \( d \) is much shorter than \( D \).

Assume the interpupillary distance \( I \) is a known constant for each observer. The
above equation implies a linear relation between the depth interval \( d \) and the binoc-
ular disparity \( \Gamma \) when the viewing distance is fixed. Namely, the larger the binoc-
ular disparity, the larger the depth interval. Therefore, when a 3D object is viewed
binocularly, the depth order of its visible vertices is available without the need of any
knowledge of the viewing distance. Prior studies have shown that people have very ac-
curate and precise depth-order judgments from stereo (Blakemore, 1970; Westheimer,
1979; Norman & Todd, 1998). This perceived depth-order information then helps the
reconstruction of relief structure (G˚ arding, Porrill, Mayhew, & Frisby, 1995; Koen-
derink, van Doorn, & Kappers, 2006), which can, at best, determine the 3D shape
up to a stretch along the line of sight, i.e., the Z-scale family (Bellhumeur, Kriegman,
& Yuille, 1999).

To identify a unique 3D shape from its Z-scale family, information other than
depth order is required. The magnitude of the depth interval between any pair of
vertices is a good candidate, since it varies systematically with the stretch along
the line of sight (i.e., Z-axis). If this depth magnitude information can be correctly
estimated and used by human observers in 3D shape perceptions, then the visible
part of a 3D shape can, in principle, be veridically reconstructed. The magnitude
of binocular disparity is useful but insufficient to precisely specify the magnitude of
a depth interval. Equation 1.1 implies that the binocular disparity \( \Gamma \) induced by a
fixed depth interval varies roughly inversely with the square of the viewing distance
Therefore, to correctly calculate depth from binocular disparity, one must scale the disparity with an accurate estimation of viewing distance according to the inverse square law.

A considerable amount of psychophysical evidence (e.g., Gilinsky, 1951; Foley, 1980; Glennerster, Rogers, & Bradshaw, 1996; Brenner & van Damme, 1999; Foley, Ribeiro-Filho, & Da Silva, 2004; Z. Li, Phillips, & Durgin, 2011; Z. Li et al., 2013; Guan & Banks, 2016) has suggested that human observers are capable of scaling the binocular disparity with their estimations of viewing distance roughly according to the inverse square law. However, their disparity scaling is not perfect due to the systematic errors in estimating the absolute value of viewing distance. And this insufficiently calibrated disparity will eventually have a systematic impact on perceiving 3D lengths as well as 3D shapes.

Prior studies have demonstrated that people perceive the lengths of in-depth intervals differently than the lengths of those intervals oriented in the frontoparallel plane. The perception of the latter is nearly accurate (except for some classical illusions, such as vertical-horizontal illusion found by Fick, 1851 and firstly reported by Wundt, 1862), whereas the perception of the former is systematically distorted. With the increase of viewing distance, an in-depth interval appears progressively shorter though its physical length remains unchanged (Gogel, 1960; Wagner, 1985; Norman, Todd, Perotti, & Tittle, 1996; Z. Li et al., 2013).

This empirical fact has important consequences for the perception of 3D metric structure, which is defined by relations among interval lengths in different 3D directions. Specifically, the ratio of the length of an in-depth interval to the length of a frontal interval (i.e., depth-to-width ratio) should be perceived with a systematic distortion such that the depth-to-width ratio of a 3D structure appears progressively smaller as this structure moves farther from the observer.
This prediction has been confirmed by a sizable body of research studying 3D shape perception from stereo using various stimulus configurations such as simple line segments (Loomis, Da Silva, Fujita, & Fukusima, 1992; Norman et al., 1996; Bradshaw, Parton, & Eagle, 1998), corrugations (Guan & Banks, 2016), cylindrical surfaces (Tittle, Todd, Perotti, & Norman, 1995; Glennerster et al., 1996), rectangles, ellipses (Thouless, 1931), ellipsoids (Johnston, 1991; Johnston, Cumming, & Landy, 1994; Brenner & van Damme, 1999), cones and pyramids (Todd & Norman, 2003). Some of these experiments used computer-generated stimuli (such as stereograms) and were conducted in reduced cue conditions (e.g., Johnston, 1991 and Tittle et al., 1995). There are also many others conducted in natural viewing environments (such as fully illuminated room and outdoors) with real objects (e.g., Thouless, 1931 and Todd & Norman, 2003). Except for a few studies reporting nearly accurate perception of 3D metric structure from stereo (Durgin et al., 1995; Frisby et al., 1996, see Todd & Norman, 2003 for a brief discussion about the potential problems in these studies), the vast majority of experiments in this area have shown the systematic distortion.

This perceptual distortion is so robust that it cannot be reversed by adding cues informing the absolute value of viewing distance, which is the only unknown factor for properly scaling the disparity (see Equation 1.1). In theory, the viewing distance can be accurately estimated by either vergence angle (Foley, 1980; Cormack, 1985), vertical disparity (Mayhew & Longuet-Higgins, 1982) or cognitive-pictorial cues such as familiar size or perspective (O’Leary & Wallach, 1980; Predebon, 1993). Empirically, however, when one or multiple of these cues have been provided to the observers of stereoscopic experiments, their performance on judging the depth-to-width ratio doesn’t improve a lot and the systematic distortion persists (Cumming, Johnston, & Parker, 1991; Sobel & Collett, 1991; Bradshaw, Glennerster, & Rogers, 1996; O’Kane & Hibbard, 2010), except for situations where the stimulus subtends a very large
visual field (at least larger than 30°, see Fox, Cormack, & Norman, 1987 and Rogers & Bradshaw, 1995 for examples).

1.2 Pizlo’s Criticisms and His Model

The vast majority of the psychophysical experiments in this field have shown that the percept of 3D structure is systematically distorted. Specifically, the depth-to-width ratio appears smaller as the viewing distance increases. This well-established and robust finding implies that a physical 3D object would appear to change shape (elongated or foreshortened) as it moves closer toward or farther away from the observer, leading to a systematic failure of metric shape constancy.

However, there remains considerable debate over this implication simply because it is against our real life experience. After all, objects don’t appear to change shape when viewed from different distances or directions. Based on this visual experience, Pizlo (Pizlo, 2008; Pizlo et al., 2014) maintains that the shape percept of an object usually is veridical (i.e., as it actually is in the world “out there”) and thereby (metric) shape constancy must exist. He criticized the prior research which didn’t find the shape constancy for their usage of impoverished stimuli. He pointed out that the stimuli used in a shape study must be well-structured and must have enough complexity, because it is these two features that uniquely separate shape from other perceptual properties of an object and thereby make shape special. He argued that these two features were “overlooked until very recently, [which] has led to a lot of confusion in the literature on shape perception” (Pizlo, 2008, p. 8).

According to Pizlo’s statement, an ideal stimulus in a shape study must be well-structured in the sense that it can be described by a series of spatial regularities. From this perspective, a crumpled piece of paper, a bent paperclip, a rock or a
potato with arbitrary curvature should not be used in a shape study at all since they are completely irregular.

Pizlo continues to make concrete statement about what kinds of spatial regularities are needed for a veridical shape recovery. These regularities include symmetry, planarity of contours and compactness.

Stemmed from these regularities, the main constraints that are essential to the recovery are defined in the following way (Pizlo, 2008). The *Symmetry constraint* says that the 3D object that is reconstructed from any perspective image of a mirror symmetrical object must be mirror symmetrical. The *Planarity constraint* says that the 3D object that is reconstructed from any perspective image of an object consisting of planar faces must also have planar faces. The *Compactness constraint* says that the reconstructed 3D object should have maximal compactness, which is defined as $V^2/S^3$ where $V$ and $S$ are the volume and total surface area of the object, among other candidates. This compactness constraint has recently been combined with a new constraint, i.e., *minimum surface constraint* defined by $1/S^3$, in order to reflect the fact that the percepts gradually change as a function of the simplicity of the 2D retinal image relative to the simplicity of the 3D interpretation (Attneave & Frost, 1969). Taking the geometric mean of the latter two constraints giving rise to the revised compactness constraint: maximum $V/S^3$ (Pizlo et al., 2014, p. 69)

Pizlo makes a further assumption that human observers have *a priori* knowledge of these constraints. Because these constraints are satisfied by most of the objects in the real world, they are somehow built in to the human mind via the evolutionary process. Hence, people are born with a very strong predilection towards these constraints. That is to say, people will always perceive a mirror symmetrical and compact 3D object with planar surfaces as long as the 2D retinal image allows it to be seen, even
though contextual cues provide no information about its symmetry, compactness and planarity (Pizlo et al., 2014).

The other key characteristic of an ideal stimulus in a shape study is the high-degree of complexity, a characteristic that makes shape quite different from all other perceptual properties (Pizlo, 2008). For example, color varies along only three dimensions: hue, brightness and saturation. Many objects “out there” will thereby have the same color. Size and weight are even simpler than color since they vary along only a single dimension. Therefore, these simple properties can not be used to effectively discriminate objects since the objects defined by them are ambiguous. Shape is unlike all of these properties in the sense that it is much more complex. For example, a simple quadrilateral shape varies along four dimensions: the ratio of lengths of two sides, plus three angles. And many objects “out there” are much more complex than quadrilaterals. Consider, for example, a human silhouette and a circumscribed circle. All of the points except those tangent points have to be moved to change a human shape to a circle and vice versa. Theoretically, the number of dimensions characterizing an arbitrary shape is infinitely large. Despite the constraints of human’s visual system, sufficient information can still be perceived. And the loss of information along one dimension (namely, the Z-axis) during the projection to retinas wouldn’t hurt shape’s complexity at all (Pizlo, 2008, pp. 1–8). Due to this high-degree of complexity, according to Pizlo, shape has the potential to sufficiently disambiguate most objects people encounter in the environment without making any use of context.

Both of the above two features are present in the truck-like stimuli used in Pizlo and his colleagues’ research (Pizlo et al., 2014, pp. 54–55). These stimuli are complex since each of them is defined by 16 vertices in 3D space (16×3=45 parameters). They are also well-structured because they are polyhedra with one plane of mirror
symmetry. Furthermore, these truck-like stimuli are ecologically relevant in the sense that they capture the basic shape of an object that people frequently perceive in everyday life, truck, with two boxes (mimicking driver’s cabin and the part that carries cargo) resting on a common plane. In order to avoid familiarity effects, all the wheels of the truck were removed and a lot of random variability was added to the 3D structures when generating the stimuli.

Tested with these truck-like stimuli, the computational model proposed by Pizlo and his colleagues can reliably and quite accurately recovered the 3D shape from a single 2D image at least when the slant of the symmetry plane is not too close to 0° or 90°, i.e., non-degenerate views, (Y. Li, Pizlo, & Steinman, 2009). Given a 2D perspective image of a randomly generated truck-like symmetrical object, the model firstly applies symmetry and planarity constraints to all the possible 3D explanations corresponding to the 2D image, leading to a restricted family (see Appendix A in Y. Li et al., 2011 for more details). Within this family, the model then looks for the candidate that satisfies the revised compactness constraint the best, namely, the 3D shape that has the maximal $V/S^3$. Finally, the model outputs this shape as the reconstructed 3D shape. At the implementation level, the above process is equivalent to solving a global optimization problem, which requires to minimize a cost function that represents the interaction between a 2D optical pattern on the retina and the aforementioned a priori constraints.

Simulation results reported in Y. Li et al. (2009), showed that the model recovered the 3D shape very well and only made errors in estimating the aspect ratio (i.e., the ratio of the extents along two orthogonal directions) of the 3D object, which is one of the 15 parameters that define the objects in the aforementioned restricted family. Moreover, the errors made in this single parameter were less than 40% for half of the recovered shapes. When the 2D image was taken from positions that
were not too close to a degenerate view (e.g., the positions where the slant of the symmetry plane was 30°, 45° and 60°), the errors made by the model were even smaller. More importantly, human observers’ recovery was highly consistent with the model’s recovery: the correlation coefficients were greater than 0.76 for all subjects.

When the binocular disparity cue is present, human observers’ recovery is even better. Using the same truck-like symmetric stimuli as before, observers’ binocular performance is almost veridical (the error is close to zero) and shape constancy is reliably achieved (Chan, Stevenson, Li, & Pizlo, 2006; Y. Li et al., 2011). To account for these results, Pizlo and his colleagues revised their original model of monocular reconstruction by taking into account the depth order information provided by binocular disparity. In their new model (Y. Li et al., 2011), the perceived depth order of vertices in a 3D shape is combined with the monocular shape constraints by means of Bayesian inference. Specifically, the probability that a recovered 3D candidate has the same depth order of vertices as the ones perceived in the stereoscopic images of the reference object is computed by a likelihood function. And the aforementioned cost function that combines all the \textit{a priori} constraints is considered as the prior probability function. Then the posterior probability function which is proportional to the product of prior and likelihood can be approximated and the candidate shape that maximizes this posterior probability is taken as the final recovered 3D shape.

This new model performed quite similar to human observers. Interestingly, both human data and model predictions showed that the accuracy of the binocular reconstructions depends on observer’s stereoacuity, namely the ability to separate two points in depth. Higher stereoacuity threshold leads to more weight assigned to monocular prior constraints, thus discounting the accuracy of binocular performance and making it more similar to the monocular performance. This finding, when combined with the satisfactory performance of the monocular model, gives rise to two
important implications. First, the depth magnitude information (namely, the exact distance between a pair of points in depth) is not necessary for a veridical 3D shape recovery from stereo whenever regularization prior constraints can be applied. Only depth order information is needed in the binocular reconstruction. And second, both monocular and binocular 3D shape percepts involve similar mechanisms that are based on a priori regularization constraints.

The second implication has gained support from several studies (Pizlo & Stevenson, 1999; Chan et al., 2006; Y. Li & Pizlo, 2011), which have been conducted to show that these regularization constraints play a critical role in the 3D shape perception. Notably, Y. Li and Pizlo (2011) found that when using the impoverished stimuli that prohibit the application of prior constraints, subjects’ performance was rather poor despite the availability of depth cues. Whereas the recovery of well-structured objects (namely the truck-like objects in this case), which allows the application of prior constraints, was quite satisfactory, even without any aid of depth cues. They thus conclude that the depth cues (e.g., binocular disparity, motion and texture), which are the main focus of the previous research in this area, plays a secondary role, at best in the 3D shape perception. The above findings can also be taken as evidence that people do have these regularization constraints in mind as prior knowledge and can effectively use them in 3D shape perception. It thereby suggests that Pizlo et al.’s model is psychologically possible.

Another important aspect related to the psychologically possibility of their model is its global optimization process. Algorithmically, their model can be regarded as a global optimization problem with a fairly large input size. According to computational complexity theory, it belongs to the group of NP-hard or NP-complete problems that are computationally intractable, because the optimal solution for this type of problems requires an exhaustively global search whose time and memory cost
grow very quickly (non-polynomially) with the size of the problem (Cormen, Leiserson, Rivest, & Stein, 2009). Therefore, it is reasonable to question whether human brain is capable of doing such a global optimization task. Indeed, finding the optimal solution by globally searching is probably beyond the brain’s power. However, computer scientists have proposed many approximating algorithms that can produce near-optimal solutions fairly quickly (within polynomial time). And psychophysical experiments have found that people can produce close-to-optimal solution to the traveling salesman problem, which is a type of intensively studied optimization problem, in time that is on average proportional to the problem size (MacGregor & Ormerod, 1996; Graham, Joshi, & Pizlo, 2000; Vickers, Butavicius, Lee, & Medvedev, 2001; Pizlo et al., 2006). It then suggests that people can approximately solve the global optimization task efficiently without performing global search (Graham et al., 2000; Pizlo et al., 2006, 2014) and Pizlo et al.’s shape perception model is, in this sense, psychologically possible.

1.3 Experimental Ideas

In the beginning of this chapter, two questions with respect to 3D shape perception have been proposed: first, what kind of additional information is used by the human brain to help the recovery of 3D shapes; and second, whether or not the recovered 3D shape is veridical. We then narrowed down these questions within the scope of 3D shape perception from stereo. Our reading of the literature in this area shows that answers to these constrained questions remain controversial. On the one hand, a substantial amount of psychophysical studies have found that 3D shape perception relies on depth magnitude information derived from stereo and 3D shape percept is systematically distorted along the Z-axis probably due to the mis-scaling of binocular disparity. On the other hand, the computational model proposed by Pizlo et al.
demonstrates that the regularization a priori constraints are the primary information required for a veridical recovery of 3D shape and depth order, rather than depth magnitude, information is needed to improve the accuracy of the recovery to a nearly perfect level. Moreover, the validity and psychological possibility of this model has gained support from human data.

Sharp contrasts exist between the two camps on 3D shape perception from stereo. These contrasts seem to result from the different stimuli and different tasks used by the two camps. To put it differently, the field of 3D shape perception from stereo was implicitly divided into several smaller subfields and the researchers in two camps chose different subfields to study and thereby found different results. In order to uncover the mechanism underlying 3D shape perception from stereo, the above two sub-theories of subfields need to be merged into a unified theory of the entire field. However, the value of filling the gap between these two camps has received little attention in the literature.

In general, the present thesis is motivated by this crucial demand of comparing and filling the gap between Pizlo et al.’s model and the findings reported by a sizable body of psychophysical research.

We start by asking the question: what leads to the gap (divergence) between the two camps? Pizlo has already proposed a candidate answer: the complexity and structure of the stimuli. Both his model simulations and human performance in his studies have shown that the complexity and structure matter a lot to the 3D shape percepts.

Aside from this point, however, there is another factor that might also lead to the gap but has largely been neglected before. Namely, the ambiguity families used by the two camps are different. The research in both camps involve the task of choosing the 3D object within a family of 3D objects to match the reference one. This family
is called the ambiguity family and it contains infinitely many 3D objects which vary along a single dimension. And all the objects in this family can project to the same retinal image according to a specific transformation.

The ambiguity family used by the vast majority of psychophysical studies in this area is the Z-scale family. The 3D objects in this family vary along the depth dimension (i.e., Z-axis). That is to say, for the objects in this family, the 3D shape of one object can change to the shapes of any other objects by an appropriate stretch or compression along the Z-axis. The Z-scale family is an essential family to be tested within 3D shape perception research since it contains the major ambiguity resulting from the information loss during the projection.

However, the Z-scale family has never been used in Pizlo et al.’s research. Instead, the 3D objects in the family they use vary along a direction that is orthogonal to the symmetry plane of the object (see Figure 1 in Y. Li et al., 2011 for an illustration). Presumably, this family bears the problem that would critically limit the generalizability of Pizlo et al.’s model. The major ambiguity of 3D shape percepts does not reside in this family, at least theoretically. Testing the ability to disambiguate shapes in the family without ambiguity raises obvious logic concerns.

When using qualified stimuli (according to Pizlo’s theory) and testing them within the Z-scale family, which of the two aforementioned frameworks would describe people’s 3D shape perception from stereo more precisely? Specifically, would the 3D shape percepts depend on depth magnitude information and be systematically distorted along the depth as found by the vast majority of prior studies? Or would Pizlo’s model still apply to this condition and thus the percepts would be veridical without the need of depth magnitude information? To address these questions, two behavioral experiments have been conducted and reported in the next two chapters.
1.4 Bayesian Inference

We employed the Bayesian inference rather than the classical frequentist inference, as appealed by John Kruschke, “It is time that we convert our research and educational practices to Bayesian data analysis” (Kruschke, 2010b). When comparing the frequentist approach to the Bayesian approach, it is not difficult to conclude that Kruschke’s appeal should be taken seriously because of the perils of the former and the merits of the latter.

Fundamentally, the way in which frequentist inference interprets probability is counter-intuitive. When people talk about probability, they usually mean the chance that a single event could happen in the future. In frequentist world, however, probability is a long-run averaged proportion. By saying the probability of event A is, for example, 0.5, the frequentist means that if we run an infinite series of Bernoulli trials then the proportion of trials where A occurs is 0.5. This counter-intuitive interpretation of probability leads to substantial misinterpretations and confusions among psychologists of some important statistical measures, like $p$ values and confidence intervals (Hubbard, 2011; Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016; Morey, Hoekstra, Rouder, & Wagenmakers, 2016).

The inference rule used in frequentist inference is also problematic: it is not based on the observed data, but depends on data that were never observed (Wagenmakers, Lee, Lodewyckx, & Iverson, 2008). This problematic inference rule itself embodies philosophical concerns. It also gives rise to practical issues. For example, null-hypothesis significance tests do not allow researchers to state evidence toward the null hypothesis and thus overstate the evidence against the null hypothesis (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

In contrast, Bayesian inference interprets probability as an individual’s subjective
degree of belief that an event will occur. The probability of event A is 0.5 means that there is 50% of chance that event A will occur in a single trial. Also, the inference rule of Bayesian approach, i.e., the Bayes rule, is built on both the prior knowledge and the observations. Therefore, the straightforward interpretation of probability and the conditional nature of Bayes rule avoid the above problems that persist in frequentist inference. Moreover, the Bayesian approach allows flexible and complete estimations of uncertainty based on its reasonable assumption that parameters are random variables rather than unknown but fixed constants as assumed in the frequentist approach. Bayesian inference can also provide the complete information of the estimated parameters whereas inferences in a frequentist framework is limited to the simple acceptance or rejection of the null hypothesis.

Graphical models (Jordan, 2004) provide an efficient way to represent complex hierarchical Bayesian model and thus serve as a powerful aid to both communication and statistical reference (Pooley, Lee, & Shankle, 2011). The nodes of a graph correspond to random variables, and the edges between these nodes correspond to the distributional or logical dependencies of the statistical model the graph represents. We adopt the notation used in recent tutorials on graphical models aimed at psychologists (M. D. Lee, 2008; Shiffrin, Lee, Kim, & Wagenmakers, 2008) in the current thesis. Square nodes represent discrete variables and circular nodes represent continuous variables. Shaded nodes represent observations and unshaded nodes represent parameters. Stochastic variables are represented by nodes with a single border and deterministic nodes are represented with double borders. Finally, independent replications of portions of the graph structure are enclosed within rectangles, which are referred to as plates.
CHAPTER 2

Experiment 1: Perceptions of the Shape of Squares

A quasi-natural environment lab was used to investigate shape perception in the current experiment. We try to avoid the experimental environment that cannot accurately reflect the structure of the real world in order to eliminate artificial confounders. A number of studies have reported that a more natural viewing condition with real objects could lead to better perceptions of shape and metric structure than the ones yielded by very reduced-cue conditions (Durgin et al., 1995; Frisby et al., 1996; Mon-Williams et al., 2000; Porrill, Duke, Taroyan, Frisby, & Buckley, 2010). And the unnatural blur gradients and conflicts between vergence and accommodation introduced by computer-generated 3D structures (like stereograms) have also been found to hinder visual perceptions and cause visual fatigue (Watt et al., 2005; Hoffman, Girshick, Akeley, & Banks, 2008).

Rather than using computer-simulated 3D stimuli, we used quasi-real 3D object. We displayed a square shape (with some distortion) on the screen of the monitor and asked the participants to adjust the shape back to a square. Since the current experiment was conducted in a well-lit environment, participants will inevitably perceive the properties of the monitor when they perceive the square that is "painted" on one of the monitor’s surface. Ideally, the mechanism that visual system uses to perceive the monitor, which is a real 3D object, should automatically transfer to the perception of this square. In this regard, we call our stimuli as quasi-real 3D object.
The slant of the square is specified by the slant of the monitor, which is manually manipulated by the experimenters before each block of trials. This design avoids the unnatural blur gradients and cue conflicts caused by the computer-simulated slanted square. It also has advantages over using real square-shape objects (like cardboards) in that it gives participants more flexible and more precise adjustments of the stimuli.

To better mimic a real-life scenario, we let participants rotate the stimulus (by pressing keys on the keyboard) within the screen plane at will as if they can observe the shape from different vantage points. Additionally, we placed several real 3D objects which people frequently encounter in everyday life, such as a stapler and a Christmas ornament within the participants’ visual field. Notice that all the objects that are available to the participants in the current experiment, like squares, the monitor and contextual objects, contain at least one axis of symmetry, a property that is essential for 3D shape perception according to Pizlo’s theory.

2.1 Methods

2.1.1 Participants

Eleven adults, including the author (YY) and the advisor (AAP), participated in the experiment at the Ohio State University. All participants had normal or corrected-to-normal vision. Except for YY and AAP, all the other participants were naïve about the purpose of the experiment and had no prior experience as participants in vision experiments.

2.1.2 Stimuli

The stimuli were computer-generated parallelograms, like the ones in Figure 2.1, each of which was divided by its two diagonals into four triangles with the two non-adjacent
triangles painted in white and the other two triangles painted in black. The shape of a displayed parallelogram was controlled by two independent factors: width-to-height ratio ($W2H$) and pose. $W2H$ was defined as the ratio of the parallelogram’s extent along the horizontal direction of the displayed screen (i.e., X-axis) to the parallelogram’s extent along the vertical direction of the displayed screen (i.e., Y-axis). Pose was defined as the degree of rotation around the stimulus center within the screen plane from the canonical position. The canonical position is the position where the two black triangles stacked along the vertical axis of the display screen and the bases of these two triangles were both parallel to the screen’s horizontal axis (like the position of the first parallelogram of each rows in Figure 2.1).

The displayed parallelograms was derived from the following two steps in sequence. First, rotate a basic square around its center within the screen plane from the canonical position by a random degree. Then stretch or compress this rotated square along the X-axis by a $W2H$ scale randomly chosen from the range of (0.5, 2). This two-step procedure ensures that the rotation of a displayed non-squared parallelogram ($W2H \neq 1$) around its center forms a non-rigid motion. For example, rotating a squeezed rectangle from its canonical position by 45° and 90° leads to two different shapes, which are a rhombus and another squeezed rectangle that is different from the original one, respectively (see Figure 2.1b). The rotation will not change the shape of the displayed parallelogram if and only if that parallelogram is a square (namely, $W2H=1$, see Figure 2.1a)

The sizes of the stimuli were differentiated into three groups, i.e., small, medium and large, according to the edge length of the basic square from which the stimuli were generated. Small, medium and large size corresponded to the edge length=5.3 cm, 8.8 cm and 12.3 cm, respectively.
Figure 2.1: An illustration of the stimuli in Experiment 1. The stimuli were computer-generated parallelograms displayed on a gray background. Each of the displayed parallelograms was generated from the following two steps in sequence. First, rotate a basic square around its center within the screen plane from the canonical position (namely, position of the first parallelogram of each rows) by a random degree. This illustration chose 45° and 90° as examples. Then stretch or compress this rotated square along the horizontal axis of the displayed screen (i.e., X-axis) by a factor randomly chosen from the range of (0.5, 2). We call this factor as width-to-height ratio (i.e., $W2H$), which is defined by the ratio of the displayed parallelogram’s extent along the X-axis to the parallelogram’s extent along the Y-axis (namely, the vertical direction of the displayed screen). The first row applied $W2H=1$ to the basic square, leading to squares with no shape distortion across different rotated positions. Whereas the second row applied $W2H=0.67$ to the basic square, leading to different shapes at different rotated positions. It thus can been seen that the two-step procedure of making the stimuli ensures that the rotation of a displayed non-squared parallelogram ($W2H\neq 1$) around its center forms a non-rigid motion.
2.1.3 Apparatus

All stimuli were online generated in MATLAB (The MathWorks, 2015) with the help of Psychophysics Toolbox Version 3.0.12 (Brainard, 1997; Pelli, 1997). Each of them was displayed on a flat-panel LCD monitor. The monitor had a visible horizontal and vertical extent of 43.08 cm × 26.92 cm, with its spatial resolution of 1680 × 1050. It was calibrated by a Minolta photometer, and the look-up table was adjusted to correct for the gamma function of the CRT. A piece of semi-transparent paper was attached nicely to the monitor without wrinkles in order to mitigate the noticeable zigzag edges of the stimulus and thus avoid an unwanted cue, which had the potential to be used in the task. Participants viewed the display in a dark room from a chin rest to restrict head movements at a distance of 100.3 cm away from the center of the screen. At this distance, one degree of visual angle spanned about 68 pixels (≈1.8 cm). Therefore, the diagonals of basic squares for small, medium and large stimuli subtend a visual angle of 4°, 7° and 10° respectively. The largest visual angle that a fully stretched stimulus can subtend is about 20°. Figure 2.2 gives an overview of the experimental layout.

2.1.4 Procedure

Participants completed a binocular session first and then completed a monocular session. Each session contained three types of blocks and the block type was determined by the viewing angle. The experimenter manipulated the viewing angle by manually slanting the upright monitor in depth about its vertical midline, by 0° (”zero”), 38.6° (”half”) or 67° (”full”) without tilt between blocks (see Figure 2.3). Therefore, the direction of X-axis, the horizontal axis of the monitor, depends on the viewing angle whereas the direction of Y-axis, the vertical axis of the monitor, stayed unchanged. Each type of block repeated twice for one subject within a session. The order of these
six blocks was either zero-half-full-zero-full-half or zero-full-half-zero-half-full and was counterbalanced across subjects.

Each block contained 12 trials: 4 trials with small stimuli, 4 trials with medium stimuli and 4 trials with large stimuli, presented in a random order. Before the start of each trial, a beep was on for 0.4 s with a blank screen to alert the participant. Then, a stimulus was displayed at the center of the monitor screen in a gray background along with the trial number and block number appeared at the bottom of the screen. We assumed that letting the participant know his/her progress during the experiment would help to motivate the participant to produce reliable performance.

The task was to make a displayed parallelogram appear as an exact square by stretching (pressing k key) or compressing (pressing j key) it along the X-axis. Recall that X-axis is the horizontal axis of the monitor and its direction depends on the
Figure 2.3: Three levels of viewing angle: zero, half and full, were used in Experiment 1. The experimenter manipulated the viewing angle before each block by manually slanting the upright monitor in depth about its vertical midline by 0° (zero), 38.6° (half) or 67° (full) without tilt.

viewing angle. Meanwhile, participants could also rotate the parallelogram around its center within the screen plane at will by turning a knob to assist the adjustments. Recall that due to the specific stimulus design in the current experiment, the rotation of a parallelogram forms a rigid motion if and only if that parallelogram is a square (see Figure 2.1). The magnitudes of the adjustments had neither upper bound nor lower bound, as long as the adjusted shapes were visible within the boundaries of the screen. Usually, a serious participant would not compress or stretch the stimulus to the extreme, i.e., shrinking to a single line or over-stretched out of the screen. Participants’ adjustments were measured by $W2H$ as defined in the Section 2.1.2. Therefore, a perfect square corresponds to $W2H=1$ while stretching a square leads to $W2H$ larger than 1 and compressing a square leads to $W2H$ smaller than 1.

Before doing the experiment, the participant was carefully instructed that there was at least three criteria that could be used to tell whether the displayed shape was a square. Namely, the four edges of a square had equal lengths and the four vertex angles of a square had equal magnitudes, the two diagonals of a square were orthogonal to each other and, in the current experiment, the rotation of a square
around its center formed a rigid motion while the rotation of other parallelograms was non-rigid. At the beginning of each session, the participant practiced with several demo trials in zero viewing angle condition where the screen plane was perpendicular to the direction of sight before he/she proceeded to the experimental trials. There was no time limit for each trial and no feedback was given. The participants were given breaks between blocks. On average, each session lasted for 40 minutes.

2.1.5 Analyses

In the analyses, we took the natural logarithms of the adjusted \(W2H\) as the dependent variable since \(W2H\) is a ratio and taking the log of it enables it to be additive, thus making the computations easier. The independent variables are viewing angle (0°, 38.6° or 67°), ocularity (binocularly viewing or monocular viewing) and the size of the stimuli (small, medium or large), each of which was repeatedly measured within subject.

2.1.5.1 Bayesian ANOVA analysis

First, we conducted a preliminary analysis using the Bayesian approach to ANOVA to select which effect(s), including main effects of each independent variable and their interaction effects, had significant impacts on participants’ performance. The idea of this analysis was introduced in (Rouder, Morey, Speckman, & Province, 2012; Rouder, Morey, Verhagen, Swagman, & Wagenmakers, 2016). Basically, a list of models are generated based on the data structure. The full model is the linear combination of the maximum number of covariates existing in the current data set. All the other alternative models can be derived from the full model by removing one or more effects. And the null model is the one that contains no effect. Then a Bayes factor is computed for each of the candidate models, including the full model and
all the alternative models, relative to the null model. The Bayes factor ($BF_{10}$) is a Bayesian model selection measure that quantifies the probability of the data under the alternative hypothesis ($H_1$) relative to the probability of the data under the null hypothesis ($H_0$). For example, $BF_{10} = 10$ indicates that the data are 10 times more likely under $H_1$ than under $H_0$. In our case, $H_0$ always corresponds to the null model and $H_1$ corresponds to a specific candidate model for which the $BF_{10}$ is computed. Since each candidate model’s $BF_{10}$ is computed relative to the same denominator, i.e., the null model, it is more useful to compare the $BF_{10}$s among candidate models than to focus on the value of each $BF_{10}$. For example, when $BF_{10}$ for model A is greater than $BF_{10}$ for model B, it is the model A rather than the model B that can better fit to the data.

In the current analysis, we formalized this idea by defining a full model, $M_f$, as

$$Y_{ijkp} = \mu + \sigma(c_i + a_j + s_k + ca_{ij} + cs_{ik} + as_{jk} + cas_{ijk} + sub_p) + \varepsilon_{ijkp}$$

(2.1)

where $Y_{ijkp}$ is the natural logarithm of the adjusted $W2H$ for the $p$th participant in the cell with $i$th level of ocularity, $j$th viewing angle and $k$th stimulus size; $\mu$ is the grand mean, $c_i$ is the effect size of the $i$th level of ocularity, $a_j$ is the effect size of the $j$th viewing angle, $s_k$ is the effect size of the $k$th stimulus size, $ca_{ij}$ is the effect size of the two-way interaction of $i$th level of ocularity and $j$th viewing angle, $cs_{ik}$ is the effect size of the two-way interaction of $i$th level of ocularity and $k$th stimulus size, $as_{jk}$ is the effect size of the two-way interaction of $j$th viewing angle and $k$th stimulus size, $cas_{ijk}$ is the effect size of the three-way interaction for the $ijk$th combination of viewing angle, ocularity and stimulus size, $\varepsilon_{ijkp}$ is a residual error term with the standard deviation $\sigma$. Note that this full model also include another factor, $sub_p$, that doesn’t correspond to any experimental manipulations. In fact, $sub_p$ is the factor that captures the participants’ variability for within-subject factors, namely ocularity, viewing angle and stimulus size in our case. This factor is also included in
the null model and thus helps the analysis to accommodate the individual differences to some degree.

The sub-models are derived by eliminating one or more effect(s) from the full model $M_f$. Thus there are $2^7 = 128$ models in total, including the full model and the null model. The Bayes factor was computed for each of the 127 candidate models using the \texttt{anovaBF} function provided by the \texttt{BayesFactor} package (Morey, Rouder, & Jamil, 2014) implemented R software. 100,000 Monte Carlo simulations were generated to compute each Bayes factor. We confirmed that the chain converged and there was no indication of autocorrelation in the samples by visually checking the trace plots of the parameters. The details of the likelihood and prior specifications can be found in Rouder et al. (2012).

2.1.5.2 Hierarchical Bayesian analysis

Although the Bayesian ANOVA analysis provides valuable information about which effect(s) contribute to the variation of participants’ performance, it has the limitation in quantifying the exact contribution of each effect. Furthermore, it is not flexible enough to correctly handle the uncertainty, especially the individual differences. From Eq 2.1, one can observe that the standard deviation, $\sigma$, doesn’t depend on individuals. That is to say, the variance of each individual is assumed to be the same. However, this assumption is easily broken in human observers.

To obtain the complete information of the estimated parameters and handle individual differences correctly while take advantage of the results derived from the Bayesian ANOVA analysis, we built a hierarchical Bayesian model that was directly driven by our intuition of the current data structure and, at the same time, took all the effects revealed in the best model selected by the Bayesian ANOVA analysis into
considerations. In this data set, we should expect variability on at least two qualitatively distinct levels to contribute to the performance, i.e., the natural logarithms of the $W2H$ adjustments. At an individual level, each participant should be expected to have differing square perceptions. For example, a shape that perceived as a square by one participant would not necessarily be perceived as a square by another participant. At a group level, participants’ performance in one condition, e.g., binocularly viewing the frontoparallel screen, should be expected, on average, to differ from their performance in another condition, e.g., monocularly viewing the oblique screen.

Figure 2.4 shows the graphical model of the current hierarchical Bayesian analysis as well as the model specifications. To implement the above intuitions about the two-level variability, our hierarchical Bayesian model divided the entire data set into $2 \times 3 = 6$ conditions and assigned each condition with a mean and a variance to capture the values of the adjustments as well as another set of mean and variance to measure the inconsistency among the adjustments. Higher group-level mean associated with the values of the adjustments infers participants, on average, are more likely to perceived the stretched (elongated) parallelograms as squares in the corresponding condition and higher group-level variance associated with the values of the adjustments infers larger variability among participants with respect to their square perceptions. Similarly, higher group-level mean associated with inconsistency implies that participants, on average, tend to make more inconsistent adjustments in the corresponding condition whereas higher group-level variance associated with inconsistency implies greater variability among participants with respect to their inconsistency. These group-level means and variances were independent across conditions and their values were fit by the data during the analysis.
Figure 2.4: Graphical model and model specifications for the hierarchical Bayesian analysis for Experiment 1. $Y_{ijk}$ referred to the natural log of the observation, adjusted $W2H$, in each trial per subject. $x_{ijk}$ indicates the stimulus size for each trial. 3, 5 and 7 correspond to small, medium and large size, separately. Since the current study has 6 experimental conditions, i.e., 2 (ocularity: binocular or monocular) $\times$ 3 (viewing angles: $0^\circ$, $38.6^\circ$ or $67^\circ$), $k$ denotes which condition the data, $Y_{ijk}$, belong to.

- **Hyper-parameters:**
  - $\mu_k$ $\sim$ Normal(0,1), $k=1,2,3,4,5,6$.
  - $\sigma_k$ $\sim$ Uniform(0,10), $k=1,2,3,4,5,6$.
  - $\mu_k$ $\sim$ Normal(-10,1), $k=1,2,3,4,5,6$.
  - $\sigma_k$ $\sim$ Uniform(0,1), $k=1,2,3,4,5,6$.
  - $\mu_k$ $\sim$ Normal(0,1), $k=1,2,3,4,5,6$.
  - $\sigma_k$ $\sim$ Uniform(0,10), $k=1,2,3,4,5,6$.

- **Parameters:**
  - $\alpha_i$ $\sim$ Normal($\mu^a$, $\sigma^a$).
  - $\theta_{ik}$ $\sim$ Normal($\mu^0_k$, $\sigma^0_k$).
  - $\lambda_k$ $\sim$ Normal($\mu^\lambda_k$, $\sigma^\lambda_k$).
  - $\sigma_k = \exp(\lambda_k)$

- **Likelihood:**
  - $\bar{Y}_{ijk} = \theta_{ik} + x_{ijk} \alpha_i$, $x_{ijk} = 3, 5, 7$
  - $Y_{ijk} \sim$ Normal($\bar{Y}_{ijk}$, $\sigma_k$)
Individual-level mean and variance of each participant’s performance in each condition were then sampled from two Gaussian distributions: one with the appropriate group-level mean and variance associated with the values of the adjustments and another one with the appropriate group-level mean and variance associated with inconsistency, respectively. Therefore, each participant had six pairs of individual-level mean and variance, each of the pairs corresponded to his/her data in one of the six conditions. Lastly, each single data point, namely the natural logarithm of an observed $W2H$ adjustment in our case, was treated as a sample that drawn from a Gaussian distribution with the appropriate individual-level mean and variance.

Given that the stimulus size effect was included in the best model selected by the previous Bayesian ANOVA analysis and the fact that stimulus size varied within each condition in this experiment, we incorporated the stimulus size effect into the current hierarchical Bayesian model by treating stimulus size as a regressor and shifting the Gaussian distribution from which a single data point was sampled by an amount equal to the product of an individual-level regression coefficient and the stimulus size associated with that data point. All the individual-level regression coefficients were sample from a single Gaussian distribution, whose mean and variance were fit by the data during the analysis.

We placed the standard normal priors for the group-level means associated with the values of the adjustments, $\mu^\theta_k$. This is a reasonable weakly-informative prior since it puts the most density around zero, which means no effect, and provides only a small amount of information for the occurrence that the effect is significantly biased from zero. For the other group-level mean parameter, $\mu^\lambda_k$, we chose the prior of Normal($-10, 1$). Note that $\mu^\lambda_k$ estimates the averaged value of the natural logarithm of the individual-level variance, $\sigma_{ik}$, rather than the $\sigma_{ik}$ itself. Conceptually, this
Normal(−10, 1) prior for a parameter on a log scale serves a similar function as a standard normal prior for the corresponding parameter on a linear scale.

We assumed a widely spread noninformative prior, Uniform(0, 10), on the group-level variance associated with the values of the adjustments, σ_k^θ, as suggested by Gelman (2006). We tested another broadly distributed prior, i.e., the inverse chi-square with one degree of freedom, recommended by Zellner and Siow (1980) and found nearly identical results. For σ_k^λ, the group-level variance associated with inconsistency, we chose a noninformative prior with smaller range, Uniform(0, 1), because of the log-scale nature of this parameter. Uniform priors with large range, like Uniform(0, 10), were also tested but MCMC chains couldn’t converge in these cases.

This graphical model, as shown in Figure 2.4, was implemented in JAGS, an open-source software package for simulation from Bayesian hierarchical models using Markov chain Monte Carlo (MCMC) (Plummer, 2003). Our results are based on two MCMC chains, each consisting of 60,000 samples collected following a burn-in period of 1,000 samples. Convergence of the chains was assessed by visually checking the trace plots of the parameters of interest.

All the eleven subjects were included in the above two analyses except two participants’ monocular data, since one (sbj110) of them did not complete the monocular session and the other participant (sbj109), when we debriefed him, reported that he used an external cue, i.e., the grid of the semi-transparent sheet attached to the screen, to facilitate his adjustments in the monocular session. Their binocular data, on the other hand, were not contaminated and thus still passed into the analyses.

### 2.2 Results

The questions we ask in this square perception experiment are twofold. First, are the square perceptions nearly veridical ($W2H$ centered around 1) or non-veridical
(W2H significantly biased from 1)? And second, how the square perceptions are influenced by other independent variables, i.e., viewing angle (zero, half or full), ocularity (binocular or monocular) and the size of the stimuli (small, medium or large).

To answer these questions, we applied two separate analyses to participants’ performance. The Bayesian ANOVA analysis provided us with the preliminary results that which effect(s), including main effects and interaction effects of the independent variables, had significant impacts on participants’ performance. We then built a more flexible hierarchical Bayesian model to acquire more precise estimations.

We have confirmed that the MCMC chains sampled in both analyses had converged. Furthermore, our preliminary analysis, the Bayesian approach to ANOVA, used a well-established model whose structure, likelihood and priors were already defined and fixed by Morey et al. (2014). And this model has also been used in many other psychological experiments (e.g., Oberauer & Eichenberger, 2013; Verhagen & Wagenmakers, 2014; Trippas, Handley, & Verde, 2014). Although the model used in our second analysis was custom-built, its structure is reasonable since it was directly driven by the intuitive nature of the current data set and its performance was found to achieve a basic level of descriptive adequacy (i.e., the ability of the model to account for and describe interesting patterns in the observed data) by posterior predictive check (e.g., Gelman, Carlin, Stern, & Rubin, 2004, pp. 165–172). Therefore, it is sensible to examine and report the results of both analyses.

2.2.1 Preliminary Results

We concluded from the Bayesian ANOVA analysis that participants’ square perceptions were influenced by all three independent variables, i.e., viewing angle, ocularity and stimulus size, as well as by the interaction of viewing angle and ocularity, because
Table 2.1: The top five models for fitting the data of Experiment 1 selected by the Bayesian ANOVA analysis according to the descending order of the Bayes factor values. The model with the greatest $BF_{10}$, i.e., the model that ranked the first, is the best.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Models</th>
<th>Bayes Factor ($BF_{10}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Angle + Ocularity + Size + Angle×Ocularity</td>
<td>$8.81 \times 10^{350} \pm 0.89%$</td>
</tr>
<tr>
<td>2</td>
<td>Angle + Ocularity + Size + Angle × Ocularity + Angle×Size</td>
<td>$1.87 \times 10^{350} \pm 1%$</td>
</tr>
<tr>
<td>3</td>
<td>Angle + Ocularity + Size + Ocularity × Size + Angle×Ocularity</td>
<td>$1.23 \times 10^{350} \pm 1.39%$</td>
</tr>
<tr>
<td>4</td>
<td>Angle + Ocularity + Size + Ocularity × Size + Angle×Ocularity + Angle×Size</td>
<td>$3.80 \times 10^{349} \pm 29.52%$</td>
</tr>
<tr>
<td>5</td>
<td>Angle + Ocularity + Size + Angle×Ocularity × Size</td>
<td>$8.75 \times 10^{348} \pm 1.24%$</td>
</tr>
</tbody>
</table>

the $BF_{10}$ associated with this model is the greatest among all the candidates. Recall that the larger the $BF_{10}$, the better the model fits the data. Table 2.1 shows the top five models ranked in descending order of their $BF_{10}$ values. Notice that the $BF_{10}$s associated with the five models are all much greater than $10^2$, indicating a *decisive* evidence that these models are much more strongly supported by the data compared with the null model, which contained no effect (see Jeffreys, 1961 for a classification scheme for the Bayes factor). More importantly, the $BF_{10}$ of the first model is about eight times greater than the second model. That means there is *substantial* evidence to support that the first model fits the data better than the second model. Therefore we are confident in taking the first model as the best one. Adding other effect(s) on this best model would lead to a model with worse performance in fitting the data.

These preliminary results, however, are conditional on the assumption of equal variance (see more discussion in Section 2.1.5.2) which might not hold in the data of the current study. The more precise estimations of the main effects of viewing angle, ocularity and stimulus size as well as the interaction effect of viewing angle and ocularity are shown in the following main results obtained from the hierarchical
Bayesian analysis. This analysis also tests the veridicality of participants’ square perceptions.

### 2.2.2 Main Results

The main results, except for consistency and stimulus size effect reported in Section 2.2.2.3, are drawn from the posterior distributions of the parameters $\mu_k^\theta$ ($k=1,2,3,4,5,6$) in the hierarchical Bayesian analysis (see Figure 2.4). As mentioned in the Section 2.1.5.2, $\mu_k^\theta$ are the group-level mean parameters associated with the values of adjustments in the six experimental conditions and higher estimated values of these parameters infer that participants, on average, are more likely to perceived the stretched (elongated along the line of sight) parallelograms as squares.

But before we go into the main results, let’s briefly review how we measured the participants’ square perceptions in the current experiment. We recorded the $W2H$ of the adjusted stimulus when the participant perceived its shape as a square and then used the natural logarithm of this value as the measurement of interest in the analysis. Recall that $W2H$ is the ratio of the stimulus’ extent along the horizontal axis of the screen divided by its extent along the vertical axis of the screen. A perfect square has its $\log(W2H)=0$, whereas stretching a square along the horizontal axis of the screen leads to $\log(W2H)>0$, and compressing a square along the same direction leads to $\log(W2H)<0$. Therefore, we will observe $\log(W2H)>0$ if participants’ mental representations of shapes are compressed along this dimension because they need to stretch more to compensate their inner compression of the shapes.

This is exactly what we found. Figure 2.5 shows the mean adjustments (measured by $\log(W2H)$ as a function of the viewing angle for binocular (the red line) and monocular (the blue line) viewing conditions, separately. The six point values correspond to the estimated means of the $\mu_k^\theta$ ($k=1,2,3,4,5,6$), averaged across samples
Figure 2.5: Posterior means of the estimated $\log(W2H)$ as a function of the viewing angle for binocular (the red line) and monocular (the blue line) viewing conditions in Experiment 1. The six point values correspond to the posterior means of the $\mu_k^\theta$ (k=1,2,3,4,5,6), the parameters of the hierarchical Bayesian analysis indicating averaged participant’s performance in each experimental condition. $\log(W2H)=0$ (denoted by the dashed line in this plot) means adjustments that produce perfect squares. $\log(W2H)>0$ means stretching a square along the horizontal axis of the screen and $\log(W2H)<0$ means compressing a square along this direction. Error bars depicted $\pm 1$ standard deviation of the posterior distribution, which is comparable to the standard error of the mean.
of their own posterior distributions. Error bars depicted ±1 standard deviation of
the posterior distribution, which is comparable to the standard error of the mean.
And the dashed line denotes the veridical perception. As is evident from the figure, 
people’ square perceptions are systematically distorted. More specifically, their men-
tal representations tended to compress the shape along the horizontal direction of 
the display screen. And this perception distortion is systematically increased as the 
viewing angle increased under both binocularly viewing and monocularly viewing, 
with monocularly viewing yields even more distortions than binocularly viewing. To 
further examine these effects, we computed contrasts from the posterior samples and 
assessed their distributions.

2.2.2.1 Ocularity Effects

Figure 2.6 shows the effect of ocularity on the mean adjustments (measured by 
log\((W2H)\)) for each viewing angle. For each MCMC sample, we computed the con-
trast of the estimated group-level mean in log\((W2H)\) of matched monocular minus 
binocular conditions under the same viewing angle. The graphs plot the posterior 
distribution of these contrasts over the entire sample space of the MCMC simula-
tion. For example, the first graph which depicting the contrast in ocularity under 
the zero viewing angle condition is the posterior distribution for \(\mu^\theta_{k=2} - \mu^\theta_{k=1}\), where 
k=1 denotes binocular viewing under zero viewing angle and k=2 denotes monocular 
viewing under zero viewing angle. The shaded areas depict the 95% highest density 
interval (HDI) of the distribution. 95% HDI is an interval that spans 95% of the 
distribution. Every value inside this interval has higher believability than any value 
outside the interval given the data and model assumptions (Kruschke, 2010a, p. 85). 
In our case, the 95% HDI at 38.6° of viewing angle is from 0.06 to 0.2 and the 95% 
HDI at 67° of viewing angle is from 0.12 to 0.36, both of which are clearly above zero
Figure 2.6: Effect of ocularity on the mean adjustments (measured by log($W2H$)) for each viewing angle in Experiment 1. The graphs plot estimated posterior probability distributions for the difference in mean adjustments between binocular and monocular viewing conditions, $\mu_{mono}^\theta - \mu_{bino}^\theta$, for each viewing angle. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution.

(the value that indicating no effect). Hence, the two 95% HDIs imply a positive effect of ocularity under oblique viewing angles, with monocular viewing giving rise to more distortions than binocular viewing. They also suggest an interaction between viewing angle and ocularity. Namely, monocular viewing causes more distortions at a more oblique viewing angle than it does when the viewing angle is less oblique. When the shapes are viewed under zero viewing angle, however, there may well be no effect of ocularity since zero is included in its 95% HDI, which goes from -0.01 to 0.05, and thus is credible.
Figure 2.7: Effect of viewing angle on the mean adjustments (measured by $\log(W2H)$) for binocular and monocular viewing conditions in Experiment 1. The graphs plot estimated posterior probability distributions for the difference in mean adjustments between full angle ($67^\circ$) case and zero angle case, $\mu_{\text{full}}^\theta - \mu_{\text{zero}}^\theta$, for each ocularity viewing condition. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution.
2.2.2.2 Viewing Angle Effects

We have found the positive ocularity effects in both obliquely viewing conditions, but not in the 0° of viewing angle condition. We then examined the effect of viewing angle under binocular and monocular viewing by computing another set of contrasts and simulating their posterior distributions. The contrasts we used here are $\mu_{k=5} - \mu_{k=1}$ for binocular viewing and $\mu_{k=6} - \mu_{k=2}$ for monocular viewing, where $k=5$ denotes binocular viewing under 67° of viewing angle, $k=1$ denotes binocular viewing under zero viewing angle, $k=6$ denotes monocular viewing under 67° of viewing angle and $k=2$ denotes monocular viewing under zero viewing angle. When computing these contrasts, we simply ignored the conditions where the viewing angle = 38.6°, since the log($W2H$)s in these conditions are between the ones in zero angle conditions and those in 67° angle conditions (see Figure 2.5). Thus, excluding these conditions from the contrasts will not affect the results.

Figure 2.7 shows the results. In both ocularity conditions, the 95% HDIs are well above zero, indicating an oblique viewing angle yielding more distortions than the zero viewing angle. Same as the results for ocularity effects, the interaction also exists here: the distortions caused by the oblique viewing angle increase as the ocularity condition changed from binocular viewing to monocular viewing.

2.2.2.3 Consistency and Stimulus Size Effect

We have found the ocularity effects, viewing angle effects and the interaction effect in square perceptions at the group level. The next question would be are these pattern consistent at the individual level? The answer is yes. The same general trends can be observed in each participant’s performance (see Figure 2.8). This consistency across participants can also be verified by the relatively small posterior means of the parameters $\sigma_k^\theta$ with $k=1,2,3,4,5,6$ (see Figure 2.4). The largest posterior mean among
Figure 2.8: Consistency in performance across different observers in Experiment 1. Ocularity effect, viewing angle effect as well as interaction effect estimated for each observer. Each colored curve represents a posterior probability distribution of the corresponding contrast for each observer. The correspondence between the colors of curves and observers’ identification numbers is shown in the legend.

The six is 0.14 with s.d.=0.05. The consistent performance within each participant is also observed and confirmed by the relatively small posterior means of the transformed parameters $\exp(\mu_k^\lambda)$ with $k=1,2,3,4,5,6$ (see Figure 2.4). The largest posterior mean among the six is 0.12 with s.d.=0.01.

Stimulus size had a trend to add a negative effect on the overall distortions since the regression coefficient parameter of the stimulus size, $\mu^\alpha$, had the 95% HDI from $-0.009$ to $-0.002$. That is to say, the smaller the stimulus is, the more mental distortion it will yield.
2.3 Conclusions and Discussion

Eleven adults adjusted the $W2H$ ratio of binocularly viewed parallelograms to make them appear as square shape (i.e., $W2H=1$) from three viewing angles (i.e., $0^\circ$, $38.6^\circ$, and $67^\circ$) under the quasi-natural viewing condition. We found that their adjustments were consistently and systematically biased when they viewed the shapes from oblique viewing angles. Namely, they overstretched the shape along the horizontal direction of the screen, the direction that has a component along the depth (Z-axis) when the viewing angle is oblique. It follows that their perceptions of the parallelograms were consistently and systematically compressed along the depth dimension at the current viewing distance ($\approx$100 cm).

Interestingly, when the shapes were displayed in the frontoparallel plane (viewing angle = $0^\circ$), participants slightly overstretched the shapes along the horizontal direction to make them appear as squares, indicating that the horizontal lengths appear shorter than the vertical lengths. This observation is not unexpected since the existence of vertical-horizontal illusion (Fick, 1851; Wundt, 1862). But this illusion has little impact on the results reported above. The results stayed almost the same after we normalized the vertical-horizontal illusion for each subject and reapplied the above analysis.

Another important finding of the current study is that this perceptual distortion varies with the viewing angle. When participants viewed the shape from a more oblique viewing angle (e.g., $67^\circ$), the overstretch along the horizontal direction of the screen was significantly greater than the one in the less oblique viewing angle condition (e.g., $38.6^\circ$). That is to say, the same shape will appear more squeezed along the line of sight as it is viewed from a more oblique viewing angle. Analogous results have also been reported in other experiments (e.g., King, Meyer, Tangney, &

To further examine the causality of the consistent and systematic distortion pattern observed in the binocularly viewing condition of the current experiment, we continued to explore the underlying mechanism. One possible mechanism is that people judge the square shape solely on the basis of its metric structure. In our case, it is simply equivalent to a length matching task or a right-angle judgement task. If they treat the current square perception task as a length matching task, then they just need to match the parallelogram’s extent along the horizontal direction of the screen, which is slanted in depth under obliquely viewing condition, to its extent along the vertical direction of the screen, which is fixed in the frontoparallel plane. If they treat the current task as a right-angle judgement task, then they only need to adjust the shape until the angle formed by the two diagonals achieve an right angle.

According to the debrief, the above mechanism seems quite possible because most of the subjects reported that their judgments mainly relied on their estimations of edge lengths and magnitudes of angles. Although we allow participants to rotate the stimulus within the screen plane at will and the rigidness of the resulting motion is also sufficient for telling whether the shape is a square or not, we found none of the eleven participants rely exclusively on this criterion and many of them did not even use this criterion for the entire experiment. If participants’ judgements in the current study do rely on their perceptions of metric structure, then our results confirm a well-established finding in this field. Namely, people’s perceptions of 3D metric structure from binocular disparity is neither accurate nor precise (as referenced in the introduction).

One of the most influential models of 3D metric structure percepts is the vector contraction model proposed by Wagner (1985). This model breaks down an extent
which is arbitrarily oriented in the 3D space into two orthogonal components, V1 and V2. V1 is the component lying in the frontoparallel plane relative to the observer (i.e., the frontal component) and V2 is the component along the observer’s line of sight (i.e., the in-depth component). This model assumes the visual frontal component, V1\textsuperscript{′}, has the same length as the physical frontal component but the visual in-depth component, V2\textsuperscript{′}, is yielded by contracting the physical in-depth component by a constant factor, c. And the perceived distance of the extent, J, is computed from the lengths of these two visual components by the following formula

\[ J = \sqrt{V1^{′2} + V2^{′2}} \]

\[ = \sqrt{V1^{2} + (cV2)^{2}} \]

\[ = \sqrt{(D \cos \phi)^{2} + (cD \sin \phi)^{2}} \]

\[ = D \sqrt{(\cos \phi)^{2} + (c \sin \phi)^{2}} \]

where D is the Euclidean distance of the extent and \( \phi \) is the angle between the norm of the extent and the line of sight.

We fit this model to the present experimental data and computed the contraction factor (c) for each binocular and oblique viewing condition (namely, binocularly viewing with viewing angle = 38.6° or 67°). In the process of computation, we took the actual edge lengths of the basic squares (i.e., 5.3 cm, 8.8 cm or 12.3 cm) and normalized them for vertical-horizontal illusion by adding the deviation in the zero viewing angle condition. These normalized lengths were then treated as the perceived distances (J) and applied to the equation 2.2. The Euclidean distances (D) were provided with the products of the measured W2H ratios and the actual edge lengths of the corresponding basic squares.

We found that the estimated value of c is significantly smaller under 67° of viewing angle (mean=0.71, s.d.=0.07) than the one under 38.6° of viewing angle (mean=0.96,
s.d.=0.09). According to the model’s setting, \( c=1 \) indicates the veridical perception with no contraction and \( 0<c<1 \) indicates a compression of the in-depth component with smaller \( c \) yielding more compression. The finding that the overall value of contraction factor is less than 1 is qualitatively consistent with the model fitting results in Wagner (1985). It implies that participants’ square shape perceptions in the current experiment depended, at least to some extent, on metric structure perception, which is systematically compressed in the depth dimension relative to the frontal dimension at a far viewing distance (e.g., Wagner, 1985; Johnston, 1991; Norman et al., 1996).

However, it is somewhat unexpected to find a significant decrease of contraction factor from a more oblique viewing angle to a less oblique viewing angle. If binocular disparity (the depth magnitude information derived from it in particular) is the only cue participants used to recover the square shapes from retinal images, then the contraction factor should be fixed in our situation where the same type of stimuli is tested using the same task with the same viewing distance under the same illuminated environment, regardless of the change in viewing angle. The shift of contraction factor at different viewing angles suggests that participants might make use of other cues besides binocular disparity.

This speculation gains further support from participants’ performance in the monocular condition. Doing the experiment with one eye completely eliminates binocular disparity cue. Judgements that exclusively governed by retinal image should yield \( \log(W2H)=0.25 \) at 38.6° of viewing angle and \( \log(W2H)=0.94 \) at 67° of viewing angle. Whereas the \( \log(W2H) \) we actually measured in the monocular condition is 0.21 (s.e.=0.03) at 38.6° of viewing angle and 0.58 (s.e.=0.05) at 67° of viewing angle (see Figure 2.5). Approximately, participants monocularly recovered the square shape to an extent of \( 0.21/0.25 = 84\% \) when the viewing angle is 38.6° and
of 0.58/0.94 ≈ 62% when the viewing angle is 67°. From this we can conclude that participants in the present study have access to some information (such as contextual cues, prior knowledge, intrinsic bias, etc) without the need of using both eyes. This information facilitates the shape recovery in both monocular and binocular condition. And the effectiveness of this information decreases as the viewing angle increases.

What is this information likely to be? The perception of stimulus’ slant seems to be a good candidate. The setting of the present experiment makes it possible to estimate the slant of the stimulus (namely, the slant of monitor screen) with only one eye. An accurate estimation of surface slant and the projected shape of the surface on the retina can, in principle, lead to a veridical percept of the shape of a slanted surface. There has been a long tradition in visual science to take slant into account when studying shape perception since Koffka (1935), but whether shape perception depends on slant perception still remain controversial. Some computational models (Foley et al., 2004; Z. Li & Durgin, 2013) that based on slant (or visual angle) percepts can fit people’s performance on metric structure judgements quite well. Whereas a dissociation between slant percept and shape percept has also been reported by some behavioral studies (e.g., Stavrianos, 1945; Bower, 1966). In the present experiment, participants might make use of the slant estimations in their square shape judgments and thereby the viewing angle effect might result from the misperception of slant or the misbinding of slant percepts and retinal images, or both. However, this explanation is highly speculative since the present experiment did not measure participants’ slant perception.

Although it is not clear what leads to the discrepancy between the judgmental errors yielded at the two different oblique viewing angles, what is clear is that participants’ perceptions of slanted squares under the quasi-natural viewing condition are systematically compressed along the depth dimension. This result is consistent with
the well-established finding in the research of 3D metric structure perception and cannot be explained by Pizlo’s model. It contradicts the prediction of Pizlo’s model at face value. The binocular percept of square shape we found in the current study is by no means close to veridicality, despite the availability of the prior constraints (3D compactness constraint might be unavailable) and tested under the quasi-natural viewing condition. Neither can Pizlo’s model provide a satisfiable explanation for the viewing angle effect. Indeed, Pizlo’s model does not always predict a nearly perfect binocular percept. Given a simple shape with just a few feature points, like the square we used in the current study, the depth order information is relatively weak and thus the model will assign more weight on the prior. In this case, binocular shape percepts will degenerate to monocular shape percepts, which are less accurate and precise. A more concrete example can be seen in the model fit to the data of subject TK in Pizlo and his colleague’s study (Y. Li et al., 2011). However, this is not the case for our study, since we do observe a significant improvement in participants’ binocular percepts compared with their monocular percepts.
CHAPTER 3

Experiment 2: Perceptions of the Symmetry of 3D Polyhedra

The stimuli we used in Experiment 1 were computer-simulated parallelograms whose slants are specified by the physical slants of the monitor. However, they might not represent the typical 3D shapes and thereby their validity of being the stimuli of a 3D shape study is under consideration. After all, the stimulus that participants directly adjust during the experiment is a 2D shape by itself. It lacks some key properties possessed by a 3D shape, like volume. Without volume, an important prior constraint in Pizlo’s model of 3D shape perception (namely, 3D compactness) cannot be applied. These slanted 2D shapes might be just as impoverished as a bunch of line segments oriented in 3D space. And the mechanism that participants’ visual system used to process them is likely to be the one used to perceive 3D metric structures, rather than the one used to perceive 3D shapes. According to Pizlo’s theory, the two mechanisms are totally different.

Indeed, the slanted 2D shapes used in Experiment 1 have more “3D flavor” than those simply slanted surfaces. They seem to be “painted” on one surface of the monitor, which is a real 3D object and can actually be seen by the participants. It is very likely that the 3D monitor will trigger the 3D shape perception mechanism and what participants actually perceive in the experiment is an integrated system of the 3D monitor and the 2D parallelogram that drawn on it. We assumed these were true, and then we assumed participants would make use of the outputs of their
perceptions of the monitor’s 3D shape in their perceptions of the parallelogram. We therefore called our stimuli as quasi-real 3D object in Experiment 1.

The above reasoning is based on a hidden assumption that the percept of a complex scene can always be explained by the percepts of the elements of the scene plus their relations. The visual system, however, might not work in this linear way. For example, the percept produced by a linear combination of a hexagon and a Y junction, which is a cube, cannot be explained by simply adding the percept of hexagon to the percept of a Y junction. An online demo of this example can be found in http://www1.psych.purdue.edu/ zpizlo/GestaltCube.

Therefore, a parallelogram drawn on a slanted monitor is probably not perceptually equivalent to a 3D object and thereby is not an ideal stimulus to be used to test 3D shape perceptions. To avoid this problem, we used Pizlo’s truck-like stimuli (namely, 3D polyhedra with one plane of mirror symmetry, firstly used in Pizlo & Stevenson, 1999) in the current experiment. This type of stimuli should be representative for the 3D objects people frequently encounter in the environment and thereby is ideal for 3D shape studies (more discussion can be found in Section 1.2).

In addition to the improvement of the stimuli, the current experiment also revised the shape perception task used in Experiment 1 to a 3D symmetry perception task. Suppose a 3D object with one plane of mirror symmetry (like the stimulus used in the current study) is placed in the 3D space such that the viewing direction (namely, the Z-axis) formed an oblique angle with its symmetry plane. Stretching or compressing this object along the Z-axis will destroy its symmetry. The resulting Z-scale family is composed of only one symmetrical 3D object and infinitely many asymmetrical 3D objects. It gives rise to the problem of interest of the current experiment: can people reliably discriminate between symmetric and asymmetric 3D objects within this Z-scale family?
Pizlo claims that 3D symmetry is an essential prior constraint for the veridical percept of a symmetrical 3D shape because people would use this prior cue to effectively constrain the possible 3D interpretations at the early stage of their shape processing. This claim has been supported by a number of studies showing that symmetry can facilitate 3D shape perception (Vetter & Poggio, 1994; Pizlo & Stevenson, 1999; Liu & Kersten, 2003; Chan et al., 2006; Treder, 2010; Y. Li & Pizlo, 2011; Y. L. Lee & Saunders, 2013). This claim strongly suggests that a symmetrical 3D shape should always be perceived as symmetrical. This straightforward implication of Pizlo’s theory, however, has not been well studied yet. Sawada (2010) has shown that people can reliably discriminate between symmetric and asymmetric 3D objects from a single 2D image, but the asymmetry tested in Sawada’s study is not created by the aforementioned Z-stretching. To our knowledge, the current study is the first one to investigate this problem within the Z-scale family.

The present 3D symmetry perception task, on the other hand, can also be done by estimating the metric structure of the displayed 3D polyhedron, without any need of the symmetry prior. Geometrically, 3D symmetry can be achieved by making the lengths of the polyhedron’s corresponding line segments equal, the magnitudes of its corresponding angles equal, or the segments connecting its corresponding vertices perpendicular to its plane of symmetry. Given this possibility, the 3D symmetry perception task used in the current experiment can be taken as a 3D metric structure perception task. A well-established finding of people’s 3D metric perception from stereo (as referenced in the introduction) is that the depth-to-width ratio of a 3D structure appears progressively smaller as the viewing distance increases. This finding should be observed in the current study if our participants do rely on metric estimations, rather than symmetry prior, to adjust 3D symmetry. In order to test this
possibility, the present experiment evaluates participants’ 3D symmetry perceptions at two different viewing distances.

3.1 Methods

3.1.1 Participants

16 adults from the Ohio State University volunteered to participate in the experiment, including the three authors (YY, AAP and JT), another two experienced psychophysical observers (XJ and MN) and 11 others who had little prior experience as subjects in psychophysical experiments. Before being tested, XJ participated in Experiment 1 and MN received some practice in other psychophysical tests. Except for the authors, all the other participants were naïve about the purpose of the experiment. All the participants had normal or corrected-to-normal vision.

3.1.2 Stimuli

Fifteen abstract 3D polyhedra were generated according to the following three-step procedure. First, a large collection of polyhedra was randomly generated, each subjected to the same constraints used in Y. Li et al. (2011). Namely, every polyhedron had 16 vertices as well as one invisible mirror symmetry plane and was composed of three convex hexahedra. The "top" hexahedron was smaller than the one in the "bottom" and the hexahedron in the middle served as the "neck" connecting the top and bottom. The back faces of these three hexahedral were coplanar and orthogonal to the symmetry plane of the polyhedron (see Figure 3.1).

Each of the polyhedra in this collection was then presented at three different viewing angles, characterized by the three slants — 30°, 45° and 60° — of the symmetry plane. Specifically, the symmetry plane of each polyhedron was first tilted by 15°
Figure 3.1: An illustration of the polyhedra stimuli used in Experiment 2. Each polyhedron is subjected to the same constraints of the stimuli used in Y. Li et al. (2011). It has one and only one invisible symmetry plane and it is mirror symmetrical with respect to this plane. The dashed line depicts the intersection of the polyhedron and its invisible symmetry plane. The polyhedron is composed of three convex hexahedra which can be seen as "top", "neck", and "bottom". The back faces, which are invisible in this figure, are coplanar and orthogonal to the symmetry plane. The front faces are the ones that corresponds to the back faces and are not coplanar. The top face refers to the "roof" of the "top" part. And the side faces refer to the ones that have no intersection with the symmetry plane. The top face, all the three front faces as well as the three side faces on one side are visible in this illustration as well as in all the polyhedra used in the experiment.
from its canonical position, then slanted at one of these three angles, and lastly tilted back by -15°. A polyhedron was in the canonical position when its center sat on the center of the screen, its symmetry plane was aligned with the line of sight (slant=90°), its top part was above its bottom part and its front faces were all facing toward the observer. Slant and tilt are the two degrees of freedom of a surface orientation in 3D space, with slant amounting to the angle between the surface normal and the line of sight, whereas tilt depicting the direction of the slant and corresponding to a rotation around the line of sight (Stevens, 1983; Rosenberg, Cowan, & Angelaki, 2013).

Finally, the fifteen polyhedra used in the current experiment were then hand-picked from the collection by the researcher so that a qualified polyhedron satisfied all of the following rules at once under each pose: 1) The top face of the object is visible, 2) All three front faces are visible and 3) All three side faces on one side of the polyhedron are visible. We used the same set of fifteen objects for all participants rather than randomly generated objects because the former allows us to aggregate the data for multiple presentations of the same object within and between subjects in the data analysis stage. Moreover, the above three rules for hand-picking objects provide observers with the most information that could be used for doing the task in the current setting and thus control the difficulties of the tasks associated with different objects at a similar and relatively low level.

The polyhedra were drawn against a gray background with visible edges rendered in black and hidden edges removed. Polka-dot textures were mapped to the faces of the polyhedra to provide more disparity cues and thus to facilitate observers’ stereo processing. The polka dots have uniform size but arbitrarily assigned colors. Note that the polka-dot texture is not continuous at the intersection between faces. Also, the size and shape of polka dots wouldn’t be distorted by the particular shape of the face they mapped to. Therefore, the texture contains no information of the shape of
Figure 3.2: Sample stimuli from the main study of Experiment 2. The graphs show two of the polyhedra used in the main study of Experiment 2. Each polyhedron is displayed at three slants (30°, 45° and 60°, from left to right respectively). These are just monocular images, but binocular images (stereoscopic images for the subject’s left and right eyes) were displayed during the experiment.

its corresponding polyhedron (see Figure 3.2). The entire set of stimuli used in the current experiment is given in the Appendix A.

There were 15 (number of polyhedra)× 3 (number of poses)=45 stimuli in total. Note that none of the three poses is close to the degenerate views (i.e., the slant of the symmetry plane is neither 0 nor 90°). The stimuli were scaled to have an average extent of 18.7 cm (S.D.=2 cm) along the X axis and an average extent of 19.5 cm (S.D.=1.5 cm) along the Y axis, subtending an average visual angle of 11° at the viewing distance of 103.5 cm and an average visual angle of 5.7° at the viewing distance of 198.8 cm. The stimuli’s extents along the Z axis were about 19 cm and
were subject to adjustments in the experiment. The directions of X, Y and Z were
defined relative to the observer: Z denotes the line of sight, the X-axis is horizontal,
and the Y-axis is vertical.

3.1.3 Apparatus

The 3D stimuli were generated online in Matlab with the help of PsychOpenGL
(http://docs.psychtoolbox.org/PsychOpenGL) and displayed binocularly using LCD
shutter glasses (NVIDIA 3D Vision 2). Two slightly different perspective images
(stereoscopic images) for subjects’ left and right eyes were computed by the parallel
lens axes algorithm, which horizontally shifts the viewpoint by some distance cor-
responding to the interocular distance (Lipton, 1991). The computations assumed
a fixed interocular distance of 64 mm for all subjects. The subject viewed a pair
of stereoscopic images through shutter glasses that were synchronized with a LCD
monitor (Acer GN246HL) driven by NVIDIA Quadro 600 graphic card so that each
eye received the image designed for this eye only. The refresh rate of the monitor was
120 Hz. Thus, the image for each eye was updated at the rate of 60 Hz, which was
fast enough to avoid flicker.

The size of the computer screen was 53.13 cm by 29.89 cm and the resolution
was 1920 pixels by 1080 pixels. The simulated viewing distance (i.e., the distance
between the viewpoint and the center of the simulated polyhedron) was the same as
the actual viewing distance, which was the distance between the subject’s eye and
the center of the monitor screen. Two different viewing distances of 103.5 cm and
198.8 cm were used in the current experiment by moving the monitor platform along
a straight track. The subject’s head was supported by a chin rest and the line of
sight was orthogonal to the monitor.
3.1.4 Procedure

Within-subject design was used. Each subject conducted two separate sessions, with one session at a viewing distance of 103.5 cm (“near”) and another session at a viewing distance of 198.8 cm (“far”). The order of the near and far sessions was counterbalanced across subjects. Each session contained two repeated blocks. Within one block, each of the 45 different stimuli was presented once in a randomized order. In each trial, one stimulus was displayed in the center of the monitor screen against a gray background. Initially, the displayed object might or might not be symmetrical, since it was produced by scaling the symmetric object along the Z axis by a random factor within the range of [0.2, 5]. Symmetry was destroyed by this scaling because the symmetry plane was slanted by 30°, 45° or 60°. Therefore, subjects’ task was to make the displayed 3D objects appear as mirror symmetrical by squeezing (pressing $j$) or stretching (pressing $k$) the objects along the Z axis. The magnitude of the adjustments were also constrained in the range of [0.2, 5]. That is to say, the participant could not squeeze the object more when the object was already squeezed by a factor of 0.2. Similarly, the participant could not stretch the object more when it was already stretched by a factor of 5.

Before doing the experiment, subjects received careful instructions telling them “Bilateral mirror symmetry means the lengths of the corresponding line segments are equal, the magnitudes of corresponding angles are equal, and the segments connecting corresponding points are perpendicular to the plane of symmetry, which is not visible.” The subject did the experiment in a quiet room where the monitor was the only source of illumination. At the beginning of each session, the subject practiced with several demo trials until he/she had the confidence to proceed to the real experiment. A colorful-polka dots image was displayed between trials for 1.5 s to mask the after effect of the stimuli. There was no time limit for the adjustments in each trial and no
feedback was given. The adjustments were expressed in the z-scaling, \( S \), the ratio of
the extent along the Z axis of an adjusted 3D object to the extend along the Z axis of
its corresponding 3D object that is in fact symmetrical. Therefore, \( S=1 \) produces a
veridical adjustment, which means the adjusted 3D object is perfectly symmetrical,
whereas deviations from 1 produce increasingly non-veridical adjustments, with \( S<1 \)
referring to over-squeezing and \( S>1 \) referring to over-stretching. 180 adjustments
were recorded for each subject: 90 adjustments in near viewing distance session and
90 adjustments in far viewing distance session. Each session took about 35 min on
average.

### 3.1.5 Analyses

The z-scaling, \( S \), of a displayed polyhedron was measured in each trial. As in the first
experiment, the natural logarithms of the measured \( S \) was taken as the dependent
variable in the analyses. Two independent variables which are of theoretical impor-
tance are the viewing angle (30°, 45° or 60°) and the viewing distance (103.5 cm and
198.8 cm).

Following the same idea as the first experiment, we first conducted a Bayesian
ANOVA analysis to get a preliminary result of which effect(s), including main effects
of each independent variable and their interaction effects, played a significant role in
affecting the participants’ \( S \) adjustments (on a log scale). Based on the preliminary
results, we next built a customized hierarchical Bayesian model to obtain the more
precise estimations of the parameters of interest, like the mean of \( S \), the mean(s) of
the effect(s) that chosen by the Bayesian ANOVA, and so on.
3.1.5.1 Bayesian ANOVA analysis

The full model used in the current Bayesian ANOVA analysis is defined as

\[ Y_{ijp} = \mu + \sigma(d_i + s_j + ds_{ij} + sub_p) + \varepsilon_{ijkp} \] (3.1)

where \( Y_{ijp} \) is the natural logarithm of the adjusted \( S \) for the \( p \)th participant in the cell with \( i \)th level of slant under the \( j \)th viewing distance; \( \mu \) is the grand mean, \( d_i \) is the effect size of the \( i \)th level of viewing distance, \( s_j \) is the effect size of the \( j \)th slant of the stimulus, \( ds_{ij} \) is the effect size of the two-way interaction of \( i \)th viewing distance and \( j \)th slant of the stimulus, \( sub_p \) denotes the variability of the \( p \)th participants after accounting for the above three effects, \( \varepsilon_{ijkp} \) is a random noise term with the standard deviation \( \sigma \).

The Bayes factor (\( BF_{10} \)) for the full model as well as for each possible sub-model, derived by eliminating one or more effect(s) from the full model, was computed using the \texttt{anovaBF} function provided by \texttt{BayesFactor} package (Morey et al., 2014) in R software. The results are based on 100 000 Monte Carlo samples for each \( BF_{10} \). We visually checked the trace plots of the parameters in the best model, the one with the largest \( BF_{10} \), and confirmed the convergence of the chain. The likelihood function and priors follow the suggestions in Rouder et al. (2012), as we did for the first experiment.

3.1.5.2 Hierarchical Bayesian analysis

As explained in the Section 2.1.5.2, the hierarchical Bayesian model we built should not only include the effect(s) selected by the preliminary Bayesian ANOVA analysis, but also capture the intuitive nature of the current data structure, namely the variabilities at the individual level as well as the group level.

Based on the preliminary results, the current hierarchical Bayesian model collapsed the data across different viewing angles and took the viewing distance as the
only independent variable along with individual differences and residual errors. The basic idea for this analysis follows the same logic as the one in the first experiment. In principle, the entire data set was split into two subsets according to their corresponding viewing distance. At the group level, the data in each subset was fit from two aspects: the values and inconsistency of $S$ adjustments across participants. Each aspect was estimated by one pair of parameters, with the mean parameter estimating the averaged amount and variance parameter indicating the individual differences with respect to the corresponding aspect.

At the individual level, each participant’s performance under each viewing distance was also summarized by a pair of mean and variance parameters. The individual-level means were sampled from a Gaussian distribution governed by the matched pair of group-level parameters associated with the values of $S$ adjustments while the individual-level variance were sampled from another Gaussian distribution governed by the matched pair of group-level parameters associated with the inconsistency. Lastly, each single data point, namely the natural log of an observed $S$ adjustment in our case, was treated as a sample that drawn from a Gaussian distribution with the corresponding individual-level mean and variance.

Figure 3.3 shows the graphical model and model specifications. Similar to the priors used in the hierarchical Bayesian model of experiment 1, weakly informative priors were placed on the group-level means and noninformative priors were placed on the group-level variances. Details about these priors can be found in Section 2.1.5.2. This model was implemented in JAGS and the results reported in the following section are based on two MCMC chains, each consisting of 9,000 samples collected following a burn-in period of 1,000 samples. Convergence of the chains was confirmed by visually checking the trace plots of all the eight hyper-parameters. No signs of auto-correlation were found.
Figure 3.3: Graphical model and model specifications for the hierarchical Bayesian analysis for Experiment 2. $y_{ijk}$ referred to the natural log of the observation, adjusted $S_i$, in each trial per subject. $k$ indicates the two viewing distances, 103.5 cm and 198.8 cm.

- **Hyper-parameters:**
  \[ \begin{align*}
  \mu_{\theta_k} &\sim \text{Normal}(0,1), \quad k=1 \text{ or } 2. \\
  \sigma_{\theta_k} &\sim \text{Uniform}(0,10), \quad k=1 \text{ or } 2. \\
  \mu_{\lambda_k} &\sim \text{Normal}(-10,1), \quad k=1 \text{ or } 2. \\
  \sigma_{\lambda_k} &\sim \text{Uniform}(0,1), \quad k=1 \text{ or } 2.
  \end{align*} \]

- **Parameters:**
  \[ \begin{align*}
  \theta_{ik} &\sim \text{Normal}(\mu_{\theta_k}, \sigma_{\theta_k}) \\
  \lambda_{ik} &\sim \text{Normal}(\mu_{\lambda_k}, \sigma_{\lambda_k}) \\
  \sigma_k &= \exp(\lambda_{ik})
  \end{align*} \]

- **Likelihood:**
  \[ y_{ijk} \sim \text{Normal}(\theta_{ik}, \sigma_k) \]
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Models</th>
<th>Bayes Factor ($BF_{10}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Viewing Distance</td>
<td>$5.60 \times 10^{19} \pm 1.13%$</td>
</tr>
<tr>
<td>2</td>
<td>Viewing Distance + Viewing Angle</td>
<td>$9.16 \times 10^{17} \pm 0.81%$</td>
</tr>
<tr>
<td>3</td>
<td>Viewing Distance + Viewing Angle + Viewing Distance $\times$ Viewing Angle</td>
<td>$1.45 \times 10^{16} \pm 1.06%$</td>
</tr>
<tr>
<td>4</td>
<td>Viewing Angle</td>
<td>$0.02 \pm 0.33%$</td>
</tr>
</tbody>
</table>

Table 3.1: The four candidate models in the Bayesian ANOVA analysis of the main study of Experiment 2. They are ranked from the best (#1) to the worst (#4) according to the descending order of the Bayes factor values.

All the 16 participants were included in the above two analyses. The outliers for each participant under each viewing distance were detected by by determining an interval spanning over the mean $\pm 3.5$ median absolute deviations (MADs), as suggested by Leys, Ley, Klein, Bernard, and Licata (2013). MAD is defined as the median of absolute deviations from the median and thus has the advantage of being very insensitive to the presence of outliers (Huber, 1981). The mean number of the detected outliers was 1.56 with s.d.=$1.16$ for each participant under each viewing distance. The detected outliers were replaced by their neighboring bound, either upper or lower bound, of the intervals computed for their corresponding cells.

### 3.2 Results

A series of two Bayesian analyses was performed with an effort to answer the following two specific questions: First, could participants veridically perceive symmetrical 3D objects as symmetrical? And second, would their perceptions depend on viewing distance and viewing angle?
3.2.1 Preliminary Results

Table 3.1 shows the results of the Bayesian ANOVA analysis. Only the viewing distance plays a significant role in affecting participants’ 3D shape perceptions. Viewing angle, that was found to significantly bias the people’s square perceptions in Experiment 1, is no longer significant here, since adding this effect to the model that only contained the viewing distance effect makes the model perform 61 times worse in fitting the data than before.

Therefore, only the effect of viewing distance as well as individual difference were considered in the hierarchical Bayesian analysis, which provides the main results reported in the next section.

3.2.2 Main Results

We want to examine the posterior estimated values of the $S$ adjustments as well as how these adjustments varied across near and far viewing distances. A perfect symmetrical polyhedron corresponds to $\log(S)=0$. Stretching a perfect symmetrical polyhedron along the depth leads to a asymmetrical shape with $\log(S)>0$, implying that participants’ mental representations of shapes are compressed along the depth. Conversely, compressing a perfect symmetrical polyhedron along the depth leads a asymmetrical shape with $\log(S)<0$, implying that participants’ mental representations of shapes are stretched along the depth.

Figure 3.4(a) shows the mean adjustments as a function of viewing distance. The two points correspond to the posterior means of two parameters depicting the group-level means of the values of adjustments ($\mu_k$, see Figure 3.3) under near and far viewing distance. Error bars depict ±1 standard deviation of the corresponding posterior distribution, which is comparable to the standard error of the mean. And the dashed line denotes the veridical perception.
Figure 3.4: Estimated log($S$) adjustments with respect to near and far viewing distances in the main study of Experiment 2. Panel a shows the posterior means of the parameters, $\mu_{\text{near}}$ and $\mu_{\text{far}}$, used in the hierarchical Bayesian analysis of the main study. These parameters estimate the values of log($S$) adjustments at the group level. Error bars depict ±1 standard deviation of the corresponding posterior distribution, which is comparable to the standard error of the mean. Panel b shows the posterior probability distributions of the above two parameters. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution. The dashed lines denote the ground truth (namely, log($S$)=0).
At the group level, we found a systematic overstretch along the depth with a tendency toward more overstretch as the viewing distance increases. It follows that the averaged perception of 3D polyhedra is systematically distorted such that the depth-to-width ratio of a polyhedron appears progressively foreshortened as it moves away from the observer.

However, unlike the clear pattern we got in Experiment 1, the above inference bears some uncertainty given the wide error bars of the posterior means. This uncertainty is more obvious when we look at the entire posterior distributions of $\mu^\theta_k$, $k=1,2$ (Figure 3.4(b)). The two 95% HDIs under near and far viewing distances are highly overlapping, indicating the nonzero difference between the perceptions yielded at the two distances are not credible. Moreover, the point that indicates no distortion, 0, is almost included in the upper HDI (the black one), suggesting the incredibility of the perceptual distortion observed in the near viewing distance.

Two possible explanations can be found for this uncertainty. First, the pattern shown by the group data is not significant by nature, implying that the perceptions might well be veridical and independent of viewing distance. The vague tendency we observed might be caused by large random errors from various sources. If this is the case, we should expect large random errors within each participant’s data.

However, we did not observe such large variability at the individual level. Instead, the adjustments within each participant are pretty consistent since the posterior means of the parameters depicting inconsistency in the adjustments of an averaged participant, namely the transformed group-level parameters $\exp(\mu^\lambda_k)$ (see Figure 3.3), are relatively small: 0.33 (s.d.=0.03) for the near viewing distance and 0.4 (s.d.=0.06) for the far viewing distance, comparable to the small inconsistency within participants found in the first experiment.
Another possible explanation is that each participant does make consistent adjustments and produces a clearly biased pattern. The patterns of adjustments across participants, however, are highly inconsistent with some qualitative difference. Averaging these qualitatively different patterns would cancel out the positive and negative effects and thereby produce a vaguely biased pattern at the group level.

We found the latter explanation more plausible in our case. Figure 3.5 shows the posterior probability distributions of estimated \( \log(S) \) adjustments in the near and far viewing distances for each participant (\( \theta_{ik} \), see Figure 3.3). All the three possibilities with respect to the viewing distance effect are observed at the individual level. Ten out of sixteen (=62.5% of) participants (denoted by orange rectangles) show a significant positive effect, revealed by the clear separation between the two HDIs with the HDI of the far viewing distance (the red one) on the right-hand side of the HDI of the near viewing distance (the black one). Statistically, no credible effect of viewing distance was found in the rest of the participants since their HDIs at the two distances are overlapped. But we can observe a slight tendency toward positive effect in two of them (denoted by gray rectangles) and a somewhat stronger tendency toward negative effect in the other four (denoted by blue rectangles). Interestingly, the ten participants who show the positive effect of viewing distance also systematically overstretch the polyhedra under both distances whereas the rest ones who do not show the positive effect either compress or nearly accurately adjust the polyhedra (except for subject #112).

The individual differences described above can be briefly summarized as follows. The majority of the participants systematically overstretched the polyhedra in depth to make them appeared symmetrical and the magnitude of the overstretch significantly increased as the viewing distance increased from about 1m to 2m. The rest of
Figure 3.5: Posterior probability distributions of estimated the log($S$) adjustments in the near (curves filled in dark gray) and far (curves filled in red) viewing distances for each observer in the main study of Experiment 2. Each curve represents the posterior probability distributions of a $\theta_{ik}$, a parameter that estimates the log($S$) adjustments for the $i$th observer in the $k$th (either near or far) viewing distance. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution. The dashed lines denote the ground truth (namely, log($S$)=0).
the participants did not stretch the polyhedra much more at the farther viewing dis-
tance. And only one participant overstretched the polyhedra to make them appeared symmetrical. The other five either overcompressed the polyhedra in depth or made nearly veridical adjustments.

The performance of the majority is highly consistent with what has been found in the studies of 3D metric structure perception from stereo. It implies that their judgments on the symmetry rely heavily on local comparisons of the lengths of corresponding line segments, the magnitudes of corresponding angles and etc. The rest of the participants, on the other hand, performed in a way that is by no means aligned with the common findings in metric perception. Rather, their performance is more or less consistent with the predictions of Pizlo’s theory in that their adjustments were not significantly and systematically influenced by varying the viewing distance and some of them can make nearly accurate judgments on 3D symmetry.

Therefore, the two qualitatively different patterns observed at the individual level might be explained by the two different strategies participants used to perceive the symmetry of 3D objects. Some of the participants used analytical strategy that based on the informative local features, while others used holistic strategy that based on the global configuration of the entire shape (like the way in which Pizlo’s computational model works). This speculation is tested by the following control experiment.

3.3 Control Experiment: Partial Shape

The control experiment is motivated by the question that whether the speculative twostategy assumption can account for the two qualitatively different adjustment patterns observed in the above main experiment. We tested this speculation by using partial polyhedra rather than full polyhedra as the stimuli. Practically, we generated exactly the same set of 15 polyhedra as the ones used in the main experiment. When
displaying these polyhedra to participants, however, we only show them the three front faces of each polyhedron (see Figure 3.6) and made the other faces invisible. Other than that, the procedure and the apparatus were exactly the same as those in the main experiment.

If the two strategies did exist and led to the distinct adjustment patterns in the main experiment, then the participants who are assumed to use the analytical strategy in the main experiment should be expected to yield a similar pattern in the control experiment as in the main experiment, because partial shape preserve
Table 3.2: The four candidate models in the Bayesian ANOVA analysis of the control study of Experiment 2. They are ranked from the best (#1) to the worst (#4) according to the descending order of the Bayes factor values.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Models</th>
<th>Bayes Factor ($BF_{10}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Viewing Distance</td>
<td>$1.50 \times 10^6 \pm 3.98%$</td>
</tr>
<tr>
<td>2</td>
<td>Viewing Distance + Viewing Angle</td>
<td>$9.28 \times 10^4 \pm 2.28%$</td>
</tr>
<tr>
<td>3</td>
<td>Viewing Distance + Viewing Angle +</td>
<td>$1.65 \times 10^2 \pm 2.63%$</td>
</tr>
<tr>
<td></td>
<td>Viewing Distance $\times$ Viewing Angle</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Viewing Angle</td>
<td>$0.006 \pm 0.48%$</td>
</tr>
</tbody>
</table>

almost all the informative local features for judging symmetry. On the other hand, the participants who are assumed to use the holistic strategy in the main experiment should be expected to perform in a significantly different pattern in the control experiment, since partial shape breaks the global configuration of the original 3D shape and thus disables the holistic strategy to a large extent. Furthermore, if the holistic and analytical strategies exhaust the strategy space of human 3D shape perception and participants are capable of using both strategies, then participants who previously used holistic strategy in the main experiment should switch to the analytical strategy in the control experiment. Therefore, participants who did not show the positive effect of viewing distance in the main experiment are expected to show, at least an obvious bias toward, this positive effect (namely more overstretch at the farther viewing distance) in the control experiment.

11 out of 16 participants of the main experiment completed the control study and their performance was analyzed in the same way as before.

At the group level, participants’ performance in the control experiment resembles the one in the main experiment. Bayesian ANOVA analysis still suggests only the effect of viewing distance and no effects of viewing angle as well as the interaction (see Table 3.2). Hierarchical Bayesian analysis found almost the same results as in the
Figure 3.7: Estimated log(S) adjustments with respect to near and far viewing distances in the control study of Experiment 2. Panel a shows the posterior means of the parameters, $\mu^\theta_{near}$ and $\mu^\theta_{far}$, used in the hierarchical Bayesian analysis of the control study. These parameters estimate the values of log(S) adjustments at the group level. Error bars depicted ±1 standard deviation of the corresponding posterior distribution, which is comparable to the standard error of the mean. Panel b shows the posterior probability distributions of the above two parameters. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution. The dashed lines denote the ground truth (namely, log(S)=0).
Table 3.3: The top four candidate models in the Bayesian ANOVA analysis of Experiment 2, combined the main study and the control study. Only the eleven participants who completed both studies were included in this analysis. The models listed here are the top four in fitting the data. The model that has the largest $BF_{10}$ (model #1) is the best one.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Models</th>
<th>Bayes Factor ($BF_{10}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Viewing Distance</td>
<td>$5.88 \times 10^{13} \pm 5.57%$</td>
</tr>
<tr>
<td>2</td>
<td>Viewing Distance + Shape Completeness</td>
<td>$1.10 \times 10^{13} \pm 61.84%$</td>
</tr>
<tr>
<td>3</td>
<td>Viewing Distance + Viewing Angle</td>
<td>$1.25 \times 10^{12} \pm 0.82%$</td>
</tr>
<tr>
<td>4</td>
<td>Viewing Distance + Shape Completeness +</td>
<td>$2.15 \times 10^{11} \pm 1.50%$</td>
</tr>
<tr>
<td></td>
<td>Viewing Distance × Shape Completeness</td>
<td></td>
</tr>
</tbody>
</table>

main experiment. A tendency toward positive effect of viewing distance was observed and the uncertainty related to this effect still remained in the control experiment at the group level (see Figure 3.7).

The observation that completeness of shapes (i.e., full shapes in the main experiment and partial shapes in the control experiment) did not induce significantly differences in participants’ performance at the group level is verified by another Bayesian ANOVA analysis. We picked all the eleven participants who completed both the main study and the control study and combined their two data sets. This combined data set was then passed to the Bayesian ANOVA analysis, whose full model contained the main effect of viewing angle, the main effect of viewing distance, the main effect of completeness of shapes as well as all the possible two-way and three-way interaction effects. As shown in Table 3.3, the best model contained only the viewing distance effect and no effect of shape completeness, indicating that no systematic effect of shape completeness was found across the main and control studies.

However, the expected shift toward the positive effect of viewing distance (namely, farther viewing distance yields larger distortion) does seem to occur when we examine the individual data. Figure 3.8 shows the individual results of the control experiment.
and was obtained and drawn in the same way as the one of the main experiment (Figure 3.5). A quick comparison between Figure 3.8 and Figure 3.5 reveals that the negative effect of viewing distance, which is found in some of participants in the main experiments (denoted by blue rectangles), no longer exists in the control experiment. Recall that, among the eleven participants who completed both the main and control experiments, six of them showed positive effect (orange group), three others showed a tendency toward negative effect (blue group) and the rest two showed almost no effect (with only a mild tendency toward positive, indicated by gray group) in the main experiment. In the control experiment, most of the participants who are previously in the orange group remain in the orange group, except for sbj#115 and sbj#117. All but one (sbj#113) participants who previously fall into the blue and gray group move to the orange group in the control experiment and sbj#113 also follows this trend by moving from blue group to gray group. Moreover, among the five participants who did not overstretch the polyhedra in the main experiment, three of them changed to overstretch the polyhedra in the control experiment while the other two (sbj#108 and sbj#109) didn’t. And those who overstretched the polyhedra in the main experiment continued to overstretch in the control, except for sbj#117.

In short, most of the participants (9 out of 11) did perform in an expected way in the control experiment. Participants who previously showed the positive effect in the main experiment and thus are assumed to be the analytical-strategy takers kept using the same strategy and showed the positive effect in the control experiment as well. Whereas those who did not show the positive effect in the main experiment and thus are assumed to be the holistic-strategy takers revealed either a clear positive effect or a tendency toward positive effect in the control experiment, suggesting a switch from the holistic strategy to the analytical strategy.

The change in performance of sbj#115 and sbj#117 is somewhat unexpected.
Figure 3.8: Posterior probability distributions of estimated the log($S$) adjustments in the near (curves filled in dark gray) and far (curves filled in red) viewing distances for each observer in the control study of Experiment 2. Each curve represents the posterior probability distributions of a $\theta_{ik}$, a parameter that estimates the log($S$) adjustments for the $i$th observer in the $k$th (either near or far) viewing distance. The shaded area under each curve depicts the 95% highest density interval (HDI) of the distribution. The dashed lines denote the ground truth (namely, log($S$)=0).
Both of them are assumed to be analytical-takers in the main experiment, but their performance in the control experiment seemed to be against the prediction of the analytical strategy and more or less consistent with the prediction of the holistic strategy. We haven’t found a good explanation for their unexpected reversal yet.

### 3.4 Conclusions and Discussion

A main experiment and a control experiment have been conducted to study people’s 3D symmetry perceptions from stereo. Participants stretched or compressed the simulated complete polyhedra (in the main experiment) or partial polyhedra (in the control experiment) along the depth dimension to make them appear as bilaterally symmetric.

The findings of the main and control experiments are summarized by the scatter-plots in Figure 3.9a and Figure 3.9b respectively. Each data point in the plot indicates the performance of one participant in the corresponding experiment, with its x-coordinate indicating the performance in near viewing distance and its y-coordinate indicating the performance in far viewing distance. The values of the point show the posterior means of the individual parameters, $\theta_{ik}$ with $k=1$ or $2$ (see Figure 3.3), which are used to estimate each participant’s log($S$) adjustments in the hierarchical Bayesian analysis (see Section 3.1.5.2). The red triangle denotes the ground truth (namely, log($S$)=0). The area above the diagonal of each scatterplot is shaded to illustrate the positive effect of viewing distance. Namely, the data point that falls into this shaded area tells the information that its corresponding participant tends to stretch the 3D shapes (either complete or partial) more along the depth than he or she did in the near viewing distance.

First of all, we found that participants can reliably adjust the symmetry of both complete and partial polyhedra. That is to say, each participant can definitely see the
Figure 3.9: Posterior means of log($S$) adjustments for each observer in the main (figure a) and control (figure b) studies of Experiment 2. Each data point in the graphs represents the estimated log($S$) adjustments for one observer, with the adjustments in the near viewing distance corresponding to its $x$-coordinate and the adjustments in the far viewing distance corresponding to its $y$-coordinate. The red triangle denotes the ground truth (namely, log($S$)=0). The area above the diagonal of each scatterplot is shaded to illustrate the positive effect of viewing distance. Namely, the data point that falls into this shaded area tells the information that its corresponding participant tends to stretch the 3D shapes (either complete or partial) more along the depth than he or she did in the near viewing distance.
change of degree of symmetry as he (or she) is pressing the key to stretch or compress a polyhedron along the depth dimension. So his (or her) adjustments indeed reflect his (or her) perception of symmetry and are quite consistent across trials. This can be seen from the relatively short error bars for each data point drawn on the scatterplots (see Figure 3.9).

This unsurprising result, however, cannot be explained and fitted by Pizlo’s computational model of 3D shape perception. Imagine his model is also a participant who is doing the present experiment. Then the input to the model is just two projection images. With the stretching along the Z-axis, the projection image itself won’t change but the spatial relation between the two images (namely, binocular disparity) will change. However, the depth order information, which is the only cue needed from the binocular disparity in this model, is not informative in the current experiment. The depth order of a polyhedron is invariant with respect to the Z-stretching. Thus, what the model see is a 3D symmetrical shape recovered from a single 2D image and this percept is independent of the Z-stretching as well as the ocularity (binocular viewing or monocular viewing).

Obviously, this is not the case for human observers. In fact, human observers’ shape percepts vary with the Z-stretching. And their monocular percepts are by no means the same as their binocular percepts. Subjects YY and AAP were asked to do the same task with only one eye. Both of them reported that they could not do the task because the adjustments yielded little change to their shape percept, which was always symmetrical. We did not report their monocular data in the Results part because their adjustments were no more than random guessing.

Despite the fact that the performance is pretty consistent within each participant, there exists considerable variance across participants. In general, participants’
performance in adjusting the 3D symmetry of complete polyhedra diverges into two
groups (denoted by the orange and blue ovals in Figure 3.9a).

The majority (10 out of 16 participants, grouped by the orange oval in Figure 3.9a)
systematically overstretched the polyhedra in depth and the magnitude of the over-
stretch significantly increased as the viewing distance increased from about 1 m to
2 m (namely, the positive effect of viewing distance). For these participants, the
binocular percepts of symmetrical polyhedra are systematically distorted. A poly-
hedron that is in fact symmetrical appears asymmetrical and the one that is in fact
asymmetrical can appear symmetrical. This result clearly contradicts the basic idea
of Pizlo’s theory, which strongly suggests that a symmetric 3D shape should always
be perceived as symmetrical. Moreover, the significant positive effect of viewing dis-
tance found in these participants is a well-known finding in people’s perceptions of
3D metric structure from stereo (as referenced in the introduction), suggesting that
their perceptions of the polyhedra’s shapes depend on their perceptions of the poly-
hedra’s metric structure. And thus the strategy they used to do the task is probably
the analytical strategy that relies on the informative local structures. Namely, mak-
ing the lengths of the polyhedron’s corresponding line segments equal, making the
magnitudes of its corresponding angles equal, or making the segments connecting its
corresponding vertices perpendicular to its plane of symmetry.

The rest of the participants (grouped by the blue oval in Figure 3.9a), on the
other hand, performed in a very different manner. Their performance is not signif-
ically influenced by the viewing distance. Instead of overstretching the polyhedra,
they either overcompressed the polyhedra along the depth or made nearly accurate
adjustments. These results, though cannot be fitted by Pizlo’s computational model
(see our previous discussion in this section), are somewhat consistent with some ideas
of Pizlo’s theory about 3D shape perception. After all, his theory does suggest veridical percepts of 3D symmetric shapes and no systematic effect of viewing distance. So we suspect that this group of participants might use some strategies that are ideologically in common with the algorithm used by his computational model. A critical idea of his model is that 3D shapes are perceived in a holistic manner. Rather than taking the local structures as the building blocks, it takes the input retinal images as a whole and finally output an integrated 3D percept. We therefore assume that this kind of holistic processing is used by the participants in this group. Somehow, they can process the input projection images as a whole. We also know that the holistic strategy they used is not exactly the same one as the holistic algorithm used by Pizlo’s computational model. But exploring the mechanism of this strategy is beyond the scope of the current study.

The control experiment was designed in a way that reduces the possibility of using the potential holistic strategy to a maximal extent while keeps all the informative local structures for judging symmetry. It is evident from Figure 3.9b that almost all the data points fall into the shaded area, which indicates the positive effect of viewing distance. Further analysis of the individual data (see Figure 3.5 and Figure 3.8) shows that the participants who previously showed the positive effect in the main experiment also showed the positive effect in the control experiment. Whereas those who did not show the positive effect in the main experiment revealed either a clear positive effect or a tendency toward positive effect in the control experiment. These results reinforce our explanations for the twofold performance observed in the main experiment. That is, most of the participants would use an analytical strategy based on the measurements of local metric structure to adjust symmetry of binocularly viewed polyhedra, while others would use a holistic strategy based on the global configuration to do the same task.
The observed wide variations in performance among different participants in the current study is not uncommon in the research of metric structure perception. Consider Todd and Norman’s (2003) study, which is representative of the studies in this field. They investigated people’s perceptions of 3D metric structure under a wide variety of conditions and found that the metric structure percept is systematically distorted and varies dramatically across subjects, stimuli, and experimental tasks. Note in particular that one of their experiments used a similar task as the one used in the present study. Todd and Norman asked their participants to adjust the depth-to-width ratio of a test pyramid, which is defined by binocular disparity, to match the shape of a target pyramid, which is a binocularly viewed cardboard object. The participants’ judgements had relatively large constant errors despite the high level of reliability of the adjustments within each participant. Todd and Norman also found a wide variation of constant errors among different participants. Some participant underestimated the depth of the target objects by 25% while another one overestimated them by 20%. These findings are highly comparable to our results. However, the negative effect of viewing distance, namely 3D shapes appear more elongated in depth as the viewing distance increases, observed in some of the observers of our main experiment has not been found in Todd and Norman’s study as well as in many other similar studies. In theory, it should not be observed in the perceptions of metric structure from stereo since it is against the pattern produced by its fundamental mechanism, namely the mis-scaling of binocular disparity. Therefore, the comparison between Todd and Norman’s (2003) study and the present study leads to an implication that, again, supports our two-strategy explanation for the twofold performance seen in participants’ adjustments of symmetry of binocularly viewed polyhedra.

Although the two-strategy explanation is quite promising according to the above discussion, it is still speculative given the small sample pool. After all, only 4 out
of 16 participants in the main experiment show the tendency toward the negative effect of viewing distance and this tendency is not statistically credible. We cannot exclude the possibility that this tendency is produced by some nongeneric strategies (or personal preferences, random noise, etc.) that only occurs under very special circumstances and cannot be generalized to other more generic contexts of 3D shape perceptions.

Unlike Experiment 1, we did not find the effect of viewing angle in Experiment 2. However, it does not necessarily follow that the two experiments tell us contradictory stories. In the first experiment, shape perception was measured by $W2H$, which is a ratio whose numerator depends on viewing angle while the denominator stays constant across different viewing angles. Hence, the effect of viewing angle on $W2H$ literally reflects how shape percepts vary with viewing angles. Whereas the current experiment measured shape perception by $S$, a ratio of the extent along the depth of a adjusted 3D object to the extent along the depth of its corresponding 3D object that is in fact symmetrical. Notice that both the numerator and denominator of this measurement depend on the viewing angle. Thus $S$ is very likely to be independent of viewing angle and the effect of viewing angle on $S$, unlike the effect of viewing angle on $W2H$ measured in the first experiment, would barely tell any useful information about the dependency of shape percepts on viewing angle. In this sense, it is not surprising to find no effect of viewing angle in the current experiment. Moreover, this negative result could help to show the validity of the three viewing angles used in the current experiment. It is known that $45^\circ$ is the farthest viewing angle from the degenerate views. We found that the other two viewing angles, $30^\circ$ and $60^\circ$, produced similar results as the ones produced by $45^\circ$. It thereby follows that the three viewing angles used in the current experiment are far from the degenerate views. The problem of degenerate views is important to be kept in mind since it would lead to the failure
of 3D shape perceptions (in both human observers’ and Pizlo’s model’s performance) and thus becomes one of the biggest concerns of Pizlo (Y. Li et al., 2011).
CHAPTER 4

General Discussion

There is considerable controversy in the literature about how human observers perceive 3D shapes from stereo. A substantial amount of psychophysical studies (e.g., Johnston, 1991; Todd & Norman, 2003; Guan & Banks, 2016) have found that 3D shape perception is built on 3D metric structure perception. And thus it requires the depth magnitude information derived from stereo to reconstruct the local structures, such as angles and line segments. These reconstructed local structures then are assembled to a 3D shape. Therefore, people’s 3D shape perception from stereo bears the same problem as their 3D metric structure perception from stereo. Namely, it is systematically distorted along the line of sight (namely, the depth dimension or Z-axis) in a way that the shape appears more foreshortened as the viewing distance increases.

However, the above opinion has been sharply contradicted by a recent theory of 3D shape perception proposed by Pizlo and his colleagues (Pizlo & Stevenson, 1999; Chan et al., 2006; Pizlo, 2008; Y. Li et al., 2009, 2011; Y. Li & Pizlo, 2011; Pizlo et al., 2014). They argue that the mechanism of 3D shape perception is totally different from the one of 3D metric structure perception because of the uniqueness of 3D shape. And they believe that the regularization constraints such as symmetry, planarity and compactness, which are absent in metric structure perception, are critical for 3D shape perception. Their computational model takes these constraints as prior cues
and recovers the 3D shape in a holistic way. Depth order information derived from stereo is needed for improving the accuracy of the recovery. With the regularization a priori constraints as well as the depth order information, this model can produce a nearly veridical 3D shape percept without any need of depth magnitude information.

We set out to explore which of these two camps can better describe people’s performance in disambiguating 3D shapes within the Z-scale family. Two behavioral experiments have been conducted under this goal.

The first experiment uses parallelograms displayed on a slanted monitor as stimuli and tests people’s ability to disambiguate the square shape from other non-squared parallelograms in its Z-scale family under a quasi-natural viewing environment. We found that participants’ shape perception is consistently and systematically distorted across viewing angles and ocularity conditions. The shape appears more foreshortened when it is more slanted away from the observer. Monocular percept is much more distorted than binocular percept and is closer to, though not the same as, the percept governed predominantly by the retinal image.

The second experiment presents stereoscopic images of 3D polyhedra with one plane of mirror symmetry (similar to Pizlo’s truck-like stimuli used in, for example, Y. Li et al., 2011) and tests people’s ability to disambiguate the symmetric polyhedron from other asymmetric ones in its Z-scale family under different viewing distances. We observed roughly two qualitatively different patterns in participants’ performance, which are probably produced by the two different strategies of doing the task. Most of the participants (10 out of 16) used the analytical strategy which relies on local comparison of metric structures. Accordingly, their shape percepts were systematically distorted such that the polyhedra appeared more foreshortened along the depth when viewed from a farther distance. On the other hand, the rest of the participants used a holistic strategy based on the global configuration. So their shape percepts were
not significantly and systematically influenced by viewing distance and some of them could make nearly accurate judgements in both near and far viewing distances. This two-strategy explanation has been reinforced by the subsequent control experiment, which was designed to reduce the possibility of using the potential holistic strategy to a maximal extent while keeping all the local structures of the original polyhedra that are informative for judging symmetry.

The above two studies investigate people’s 3D shape perception from stereo by examining their ability to disambiguate 3D shapes within the Z-scale family. They were conducted using different stimuli and different tasks under different viewing environment. We found that the opinion that 3D shape percept depends on 3D metric structure percept can account for participants’ performance across the two studies to a large extent. Almost all the participants’ performance in the square study and most of participants’ performance in the symmetry study is in accordance with this prediction.

On the other hand, Pizlo’s computational model fails to account for the results of both studies. His model of 3D shape perception from stereo makes use of the depth order information derived from binocular disparity and does not involve the depth magnitude information at all. When stretching or compressing a 3D shape along the Z-axis, depth magnitude information of this shape changes systematically whereas its depth order is always the same. That is to say, this model is blind to the changes produced by the Z-stretching and the shapes within a Z-scale family all look the same to it. Obviously, this is not the case for human observers. Participants in both studies can make adjustments with high degree of reliability. It shows that people can definitely see the change in shape when this shape is undergoing a Z-stretching.

However, Pizlo’s arguments about the uniqueness of shape and the holistic fashion
of 3D shape processing do gain some support from the present research. The comparison between the first and second studies shows that some performance that cannot be explained by metric structure perception but is more aligned with the prediction of Pizlo’s theory emerges when the stimulus carries more features of a typical 3D shape and is tested in a task that puts more emphasis on the shape itself.
References


Psychophysics, 68(7), 1124–1139.


Kruschke, J. K. (2010b). *An open letter to editors of journals, chairs of departments,
directors of funding programs, directors of graduate training, reviewers of grants and manuscripts, researchers, teachers, and students. Retrieved from http://www.indiana.edu/~kruschke/AnOpenLetter.htm


bayes factors for common designs. *R package version 0.9, 8.* Retrieved from
http://bayesfactorpcl.r-forge.r-project.org/

for rotating objects undergoing affine stretching transformations. *Perception &

ordinal depth relations on smooth surfaces and in empty space. *Perception,

shape from specular highlights, deformations of shading, and other types of
visual information. *Psychological Science, 15*(8), 565–570.

perception of three-dimensional length. *Journal of Experimental Psychology:
Human Perception and Performance, 22*(1), 173–186.

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when
more features must be remembered for each object. *Memory & Cognition,
41*(8), 1212–1227.

O’Kane, L. M., & Hibbard, P. B. (2010). Contextual effects on perceived three-

cues in stereoscopic depth constancy. *Perception & Psychophysics, 27*, 131–
135.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transform-

3145–3161.


Stavrianos, B. K. (1945). The relation of shape perception to explicit judgments of inclination. *Archives of Psychology, 296*.


Appendix A: The Stimuli of Experiment 2

Figure A.1: The entire set of stimuli used in the main study of Experiment 2. These are just monocular images, but binocular images (stereoscopic images for the subject’s left and right eyes) were displayed during the experiment.