

Article

Theoretical Treatment of Limitations Inherent in Simple 3D Stimuli: Triangles and the P3P Problem

Vasiliy Minkov [†] and Tadamasa Sawada ^{*ID}

School of Psychology, HSE University, 101000 Moscow, Russia; proveyourselfmail@gmail.com

^{*} Correspondence: tada.masa.sawada@gmail.com or tsawada@hse.ru[†] Present address: The Institute for Information Transmission Problems (Kharkevich Institute), Russian Academy of Sciences, 127051 Moscow, Russia.

Abstract: Understanding the visual stimulus in a psychophysical experiment, theoretically, is critical for controlling the experiment, for interpreting the empirical results of the experiment, and for discussing the mechanisms the visual system used to get these results. This fact encourages visual scientists to use “simple” visual stimuli in their experiments. A triangle is one of the simplest stimuli that has been used by psychophysicists to study 3D perception. It has also been used to compose the polygonal meshes that represent complex 3D surfaces in computer graphics. The relationship between the shape, orientation, and retinal image of a triangle has also been studied as the Perspective-3-Point problem (P3P). In this study, the statistical properties of this relationship between the 2D retinal image of a triangle and its recovered 3D orientation were tested in a simulation experiment whose results showed that a triangle is qualitatively different from more complex shapes that have been used to recover 3D information from their retinal images. This raises an important question, namely, how many, if any, inferences about our visual system can be generalized to our perceptions in everyday life when they are based on psychophysical experiments that used very simple visual stimuli such as triangles.

Keywords: 3D perception; depth perception; P3P problem; shape constancy; shape ambiguity; visual space



Citation: Minkov, V.; Sawada, T. Theoretical Treatment of Limitations Inherent in Simple 3D Stimuli: Triangles and the P3P Problem. *Vision* **2021**, *5*, 10. <https://doi.org/10.3390/vision5010010>

Received: 18 October 2020
Accepted: 9 February 2021
Published: 17 February 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Understanding the visual stimulus in a psychophysical experiment, theoretically, is critical for controlling the experiment, for interpreting the empirical results of the experiment, and for discussing the mechanisms the visual system uses to get these results. This goal has encouraged visual scientists to use “simple” visual stimuli in their experiments. A 2D retinal image can, theoretically, be decomposed into points, contours, gratings, and Gabor patterns. The perception of such stimuli has also been studied in psychophysical experiments (e.g., [1,2]). Our theoretical understanding of these simple elements taken with the empirical studies of their perception permits us to discuss, systematically, the perception of more complex stimuli that are composed of such elements.

The perception of 3D information cannot be studied in the same way that the perception of 2D information on the frontoparallel plane can be studied because projecting a 3D scene on to a 2D retina is a *well-posed forward* problem but recovering this 3D information in the scene from a 2D retinal image is an *ill-posed inverse* problem [3,4]. There are infinitely many possible 3D interpretations of the 2D retinal image. The visual system can resolve this problem by using a priori constraints in the 3D scene, such as mirror-symmetry and volume [4–7]. The visual system can also use a number of depth cues for perceiving the 3D information [8]. Consider, for example, that human beings normally view a 3D scene with a pair of eyes. These eyes are separated about 6.5 cm so their retinal images of the scene will be slightly different from one another. This difference between the stereo-retinal images is called “binocular disparity”. The visual system can use this cue to perceive depth.

Can a 3D scene or a 3D object be decomposed into its simple parts in the same way that its 2D retinal image has been decomposed? The visual system could segment objects in the scene and process them individually [9–11]. These individual objects can even be segmented into smaller components, making it possible to recognize them [12–14]. The relationship between these components and their 2D retinal images can be characterized by what has been called their “non-accidental properties” [6,14,15]. These non-accidental properties are invariant features in the images of the components, and these image features can be assumed to play an important role for perceiving 3D information from the 2D retinal image.

Often, there is even more reduction of the visual stimuli employed in computer vision and in computer graphics where 3D scenes and 3D objects are often represented as compositions made up of points, or of polygons. These polygons are used to compose a polygonal mesh that represents, approximately, any complex surface of a scene and of an object. Triangles are commonly used to compose the polygonal mesh because a triangle is the *simplest* polygon that can enclose a surface that is always planar. Note that these triangles and dots are two of the simplest stimuli that have been used to study 3D perception. It has been shown that the 3D perception of these triangles, and dots, as well as some other very simple visual stimuli is *not* veridical while the perception of 3D scenes and objects “out there” is veridical in our everyday life (see References [3,4] for discussions). This difference in the veridicality of the 3D perception can be attributed to the geometrical properties of the stimuli employed. The visual system uses a priori constraints and depth cues to establish veridical 3D perception. Theoretically, both the constraints and the cues require visual stimuli that have at least some geometrical complexity. So, an understanding of the geometrical properties of the stimuli employed is necessary if one wants to understand the difference observed between veridical and non-veridical 3D perception (see References [16–18] for a discussion about the importance of Theories in Science).

In this study, we analyzed numerically: (i) the geometrical properties of the relationship between a triangle in a 3D scene and its 2D retinal image, and (ii) the retinal images of triangles that were used as visual stimuli in two prior psychophysical studies. These triangles were discussed in detail in this study because these triangles can be regarded as the kind of elements that can be used to compose a 3D scene. They are the simplest polygons that can enclose planar surfaces, and they can even represent, approximately, a more complex surface by using them to compose a polygonal mesh.

2. Analysis

The relationship between a triangle ABC in a 3D scene and its 2D perspective projection to a retina can be represented by the tetrahedron $EABC$ shown in Figure 1. The bottom face of the tetrahedron is the triangle ABC and the apex E represents the center of projection in an eye. The retinal image of ABC can be represented by three visual angles θ_{BC} , θ_{CA} , and θ_{AB} at E . The shape of the triangle ABC can be characterized by two angles ω_A and ω_B at the vertices A and B . The third angle ω_C of the triangle ABC is $\omega_C = 180^\circ - \omega_A - \omega_B$. The size of ABC can be controlled by the length of the line-segment AB . The length of AB ($\|AB\|$) can be set to 1 without any loss of generality. If the size of ABC changes by a factor of s ($\|AB\| = s$), the size of the tetrahedron $EABC$ changes by a factor of s while all of the angles of $EABC$ remain constant. Note that the distance and size of ABC from the center of projection E changes by a factor of s while the orientation of ABC is unchanged.

First, consider recovering the shape of the triangle ABC from its retinal image when both the orientation of ABC in a 3D scene and the retinal image are given. The shape can be uniquely determined by finding the intersection of the lines of projection with a plane that has the given orientation. The orientation can be arbitrary unless the normal to the plane is perpendicular to any of the lines of the projection. The distance of the plane from the center of projection E characterizes the size of ABC .

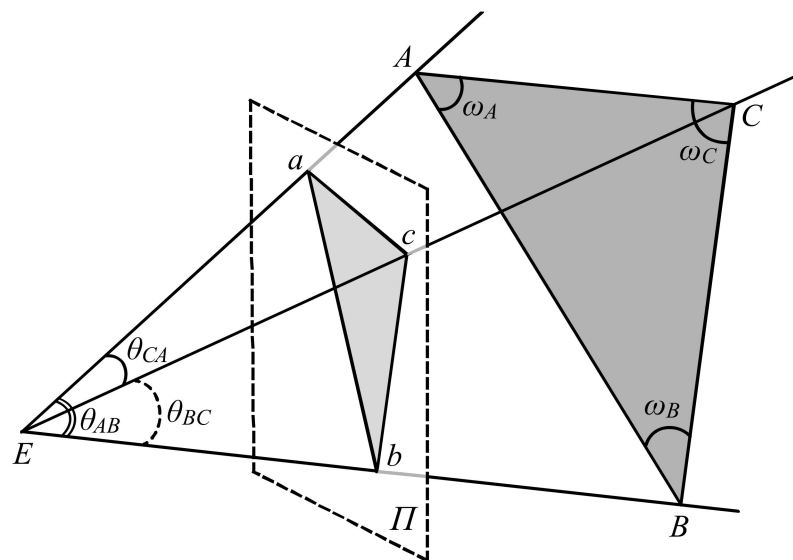


Figure 1. A perspective projection from the triangle ABC in a 3D scene to the triangle abc in the 2D image plane Π from the center of projection E . This projection can be represented as the tetrahedron $EABC$.

Next, consider recovering the orientation of the triangle ABC from its retinal image when the shape of ABC and the retinal image are given. This problem is relevant with, for example, shape recognition, shape reconstruction, and mental rotation tasks. It is known as the Perspective-3-Point (P3P) problem [19–28]. It has been proven that there are 0, or up to 4, possible 3D interpretations of the triangle for the given shape ($\omega_A, \omega_B, \omega_C$) as well as for the size s and for the retinal image of the triangle given ($\theta_{BC}, \theta_{CA}, \theta_{AB}$). These interpretations correspond with the solutions of the quartic equation that is used to solve the P3P problem (see Reference [28]).

The P3P problem in our study was formulated as the relationship between the triangle ABC and the visual angles θ_{BC}, θ_{CA} , and θ_{AB} [20–25]. This allows us to control the retinal image of the triangle ABC with only 3 parameters ($\theta_{BC}, \theta_{CA}, \theta_{AB}$). The recovered orientation of the triangle ABC is characterized by the distance l_A, l_B , and l_C of the vertices A, B , and C from the center of projection E . The recovered positions of A, B , and C can be written as:

$$\begin{cases} A = l_A V_A \\ B = l_B V_B \\ C = l_C V_C \end{cases} \quad (1)$$

where V_A, V_B , and V_C are unit vectors representing the lines of sight from E to A, B , and C . The vectors V_A, V_B , and V_C can be derived from the retinal image of the triangle ($\theta_{BC}, \theta_{CA}, \theta_{AB}$). The distance l_A, l_B , and l_C are restricted to be positive so that the recovered positions of A, B , and C do not extend behind the center of projection E .

Note that the P3P problem in computer vision [19,27,28] is formulated as a relationship between ABC and its planar perspective image (the triangle abc on the image plane Π in Figure 1) in a calibrated camera. The visual angles θ_{BC}, θ_{CA} , and θ_{AB} can be computed from this calibrated image.

2.1. Monte-Carlo Simulation

We tested the frequencies of the number of possible 3D interpretations of the triangle ABC for the retinal image in two Monte-Carlo simulation experiments by using an algorithm developed by Fischler & Bolles [21] for solving the P3P problem. The shapes of the triangle ($\omega_A, \omega_B, \omega_C$) and the retinal image ($\theta_{BC}, \theta_{CA}, \theta_{AB}$) were randomly generated in each trial by randomly sampling $\omega_A, \omega_B, \theta_{BC}, \theta_{CA}$, and θ_{AB} from uniform distributions. The sampled variables of $\omega_A, \omega_B, \theta_{BC}, \theta_{CA}$, and θ_{AB} were independent from one another but were

constrained so that $\omega_A, \omega_B,$ and ω_C form the triangle and $\theta_{BC}, \theta_{CA},$ and θ_{AB} form an apex of the tetrahedron: $\omega_A + \omega_B + \omega_C = 180^\circ, \theta_{BC} + \theta_{CA} + \theta_{AB} < 360^\circ, \theta_{BC} + \theta_{CA} > \theta_{AB}, \theta_{CA} + \theta_{AB} > \theta_{BC}, \theta_{AB} + \theta_{BC} > \theta_{CA}$. Additionally, the shape of the triangle was restricted by an additional constraint, namely, $10^\circ < \omega_A, \omega_B, \omega_C < 170^\circ$. With this done, the possible 3D interpretations of the triangle for the retinal images $(\theta_{BC}, \theta_{CA}, \theta_{AB})$ are computed by using an algorithm developed by Fischler & Bolles [21] for solving the P3P problem. This algorithm was implemented in a C++ program. We confirmed that this program is both more reliable and faster than other existing programs that have been used for the P3P problem (Appendix A).

In the first experiment, the ranges of the sampling of $\theta_{BC}, \theta_{CA},$ and θ_{AB} were set to $0.1^\circ < \theta_{BC}, \theta_{CA}, \theta_{AB} < \theta_{\max}$, where θ_{\max} is an independent variable ($2^\circ, 4^\circ, \dots, 118^\circ, 120^\circ$). There were 4×10^8 trials for each value of θ_{\max} . In the second experiment, the ranges of sampling were set to $\theta_{\max}/2 < \theta_{BC}, \theta_{CA}, \theta_{AB} < \theta_{\max}$.

The results of this simulation are shown in Figure 2. The ordinates show the frequencies of the numbers of possible 3D interpretations. The abscissa shows θ_{\max} that controls the range of the sampling. The four curves show the numbers of possible 3D interpretations.

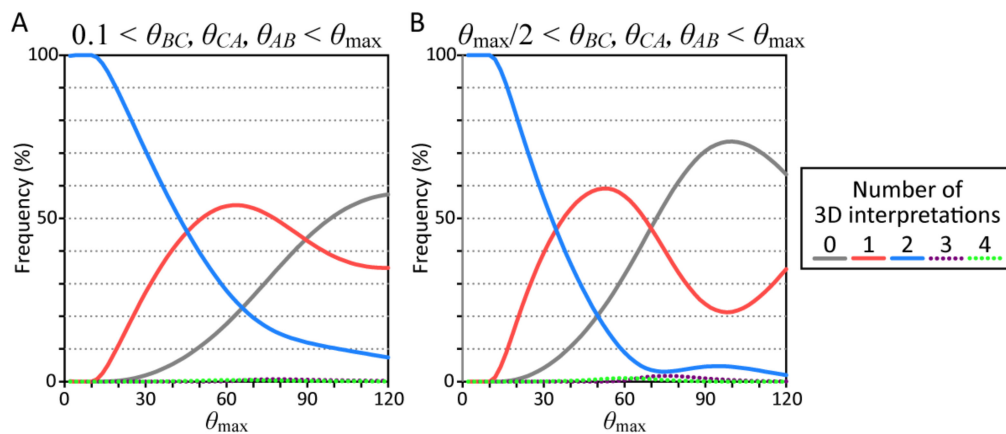


Figure 2. Results of two Monte-Carlo simulation experiments. The ordinate shows the frequency of the numbers of possible 3D interpretations and the abscissa shows θ_{\max} . The five curves show the numbers of possible 3D interpretations. (A) The visual angles $\theta_{BC}, \theta_{CA},$ and θ_{AB} were sampled between 0.1° and θ_{\max} . (B) The visual angles $\theta_{BC}, \theta_{CA},$ and θ_{AB} were sampled between $\theta_{\max}/2$ and θ_{\max} .

These results show the frequency of obtaining two possible interpretations is almost 100% (>95%) if the visual angles $\theta_{BC}, \theta_{CA},$ and θ_{AB} are small ($\theta_{\max} \leq 14^\circ$ in Figure 2A,B). We also found that the frequency of 2 possible interpretations decreases as the retinal image becomes larger. The number of possible interpretations is often 0 (>60%) if all the visual angles $\theta_{BC}, \theta_{CA},$ and θ_{AB} are larger than 40° ($\theta_{\max} \geq 80^\circ$ in Figure 2B). This number is rarely 0 (<1%) if all of the visual angles $\theta_{BC}, \theta_{CA},$ and θ_{AB} are smaller than 20° ($\theta_{\max} \leq 20^\circ$ in Figure 2A,B). The number of possible interpretations is rarely three or four for any value of θ_{\max} .

Note that the projection from the triangle to its retinal image is *perspective* but it can also be approximated well with an *orthographic* projection with uniform scaling when the visual angle of the triangle is small. The two possible interpretations of the small retinal image ($\theta_{\max} \leq 14^\circ$ in Figure 2A,B) are analogous to the depth reversal ambiguity of an orthographic image of a 3D wire-frame object such as a Necker cube [29]. We confirmed that the orientations of the triangle in the two possible interpretations are approximately depth reversals of each other.

The number of possible interpretations of the triangle is almost always 1 or more than 1 if the image is sufficiently small ($\theta_{\max} \leq 20^\circ$ in Figure 2A,B) but is often 0 if the image is sufficiently large ($\theta_{\max} \geq 80^\circ$ in Figure 2B). These trends were examined by performing an additional analysis of the effect of small and large retinal images: $(\theta_{BC}, \theta_{CA}, \theta_{AB}) = (9^\circ,$

10°, 11°) and (90°, 100°, 110°). The shapes of the triangles that could be projected to these images were computed by using the same program we used to solve the P3P problem. The results of these analyses are shown in Figure 3. Each point in these maps represents the shape of a triangle ABC. The abscissas and the ordinates show two angles ω_A and ω_B of the triangle ABC. The third angle ω_C was computed as $\omega_C = 180^\circ - \omega_A - \omega_B$. The colors of the point indicate the number of possible interpretations of the triangle with a specified shape ($\omega_A, \omega_B, \omega_C$). These trends were also observed in the individual retinal images. Almost any triangular shape ($10^\circ < \omega_A, \omega_B, \omega_C < 170^\circ$) can be projected to a small retinal image and there were two possible orientations for many of the individual shapes. With the large retinal images, less than half of the triangular shapes can be projected, so large retinal image restricts the shape of the triangle. This suggests that a large retinal image of a triangle can, to some extent, actually serve as a cue for the shape of the triangle.

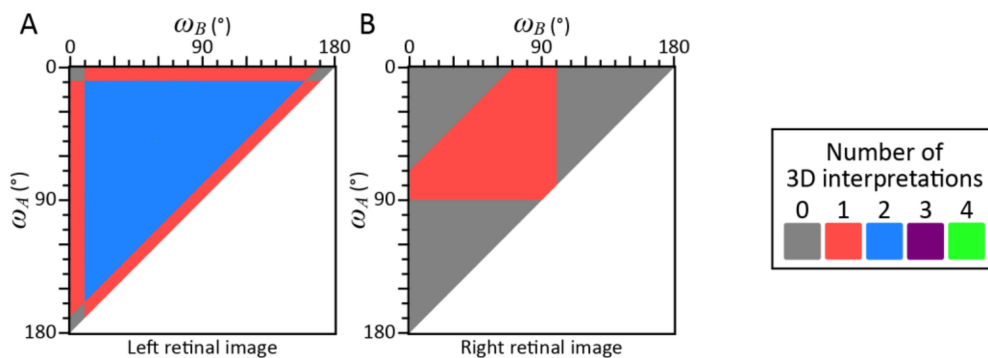


Figure 3. Results of the analyses of two retinal images of triangles where $(\theta_{BC}, \theta_{CA}, \theta_{AB}) = (10^\circ, 15^\circ, 20^\circ)$ in (A) and $(90^\circ, 100^\circ, 110^\circ)$ in (B). The ordinate and abscissa show two angles ω_A and ω_B of the triangle ABC. The colors indicate the number of possible interpretations of the triangle. White regions indicate where the shapes of the triangle would not be valid.

2.2. Analyzing the Retinal Images of Triangles That Have Served as Visual Stimuli

We began by examining the shape of a triangle that was projected to a specified retinal image. We analyzed the retinal images of the triangles that had been used in (i) Beck & Gibson’s [30] Experiment 1, and in (ii) Watanabe’s [31] Condition 3, where they studied (i) the relationship between the perceived shape of the triangle and its orientation in a 3D scene and (ii) the distortion of a perceived 3D space by comparing the visual stimuli with their observers’ responses. Note that the Watanabe’s [31] paper provides important support for Indow’s [32] theory that perceived space is distorted and that this distortion is hyperbolic. We chose these studies because of the clarity of the authors’ descriptions of the visual stimuli they used in their experiments and the simplicity of their stimuli. In both of these experiments, the observers were shown the triangles in dark rooms and they responded (i) by constructing its shape and (ii) by adjusting the positions of its vertices as well as the positions of a few added points. The triangles were viewed monocularly in Beck & Gibson and binocularly in Watanabe. The shapes of the triangles that could be projected to retinal images were computed by using the program we used to solve the P3P problem.

The results of these analyses are shown in Figure 4. Each point in these maps represents the shape of the triangle ABC. The abscissas and the ordinates show two angles ω_A and ω_B of the triangle ABC. The third angle ω_C could be computed as $\omega_C = 180^\circ - \omega_A - \omega_B$. Colors of the point indicate the number of possible interpretations of a triangle with a specified shape ($\omega_A, \omega_B, \omega_C$).

The three panels of Figure 4A show the number of possible interpretations for the retinal images of the triangle that had 3 different orientations in Beck & Gibson [29]: $(\theta_{BC}, \theta_{CA}, \theta_{AB}) = (5.538^\circ, 5.538^\circ, 6.573^\circ)$, $(4.928^\circ, 4.928^\circ, 6.638^\circ)$, and $(4.222^\circ, 4.222^\circ, 6.689^\circ)$. These images are small ($\theta_{BC}, \theta_{CA}, \theta_{AB} < 10^\circ$), and they are consistent with almost any shape of the triangle. Moreover, note that usually there were only two possible interpretations for each shape and that the number of possible interpretations is never 3 or 4.

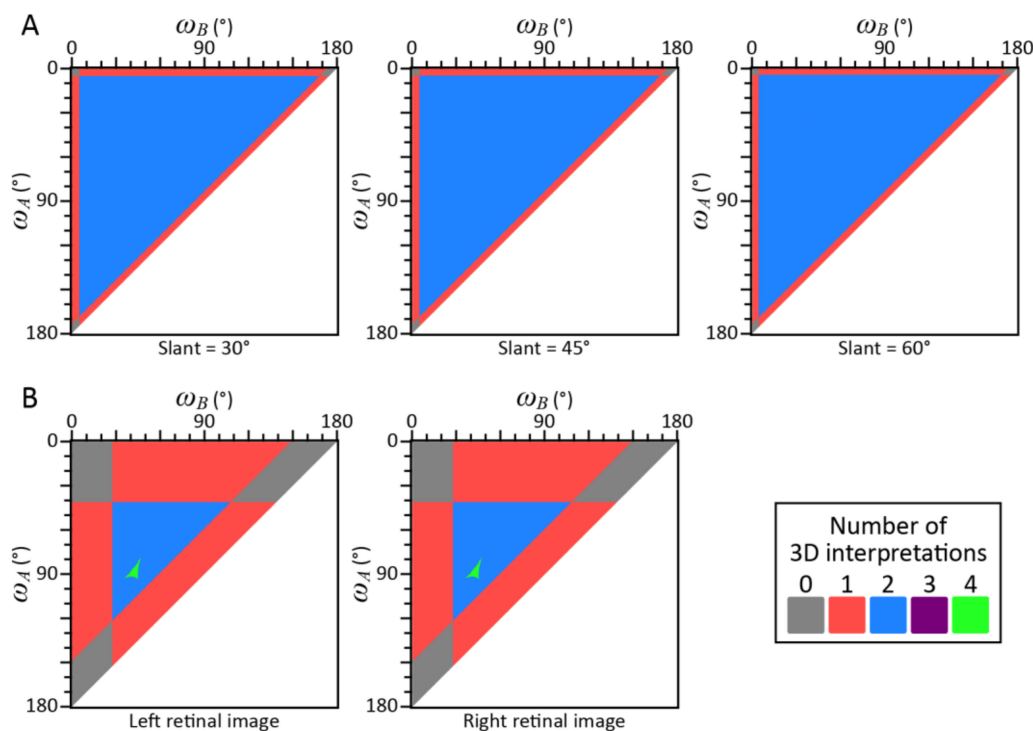


Figure 4. Results of the analyses of the retinal images of triangles: in (A) Beck & Gibson [30] and in (B) Watanabe [31]. The ordinate and abscissa show two angles ω_A and ω_B of the triangle ABC. The colors indicate the number of possible interpretations of the triangle. White regions indicate where the shapes of the triangle would not be valid. (A) The results of the analysis of the retinal images of the triangle that had 3 different orientations in Beck & Gibson [30]: $(\theta_{BC}, \theta_{CA}, \theta_{AB}) = (5.538^\circ, 5.538^\circ, 6.573^\circ), (4.928^\circ, 4.928^\circ, 6.638^\circ),$ and $(4.222^\circ, 4.222^\circ, 6.689^\circ)$. (B) The results of the analysis of the retinal images in Condition 3 of Watanabe [30] with binocular viewing. His retinal images $(\theta_{BC}, \theta_{CA}, \theta_{AB})$ were $(41.19^\circ, 27.68^\circ, 31.01^\circ)$ for the left eye and $(41.18^\circ, 27.78^\circ, 30.91^\circ)$ for the right eye.

The left and right panels of Figure 4B show the number of possible interpretations of the left and right retinal images in Watanabe’s [31] Condition 3. The individual images cannot be projected from about 20% of the triangular shapes. The number of possible interpretations was often 1 or 2 and it was rarely 4. The number 3 was not observed. This ambiguity remained even when the test was done binocularly. Many triangular shapes can be projected to both of the retinal images.

The geometrical ambiguity of the visual stimuli shown in this analysis can explain the empirical results in Beck & Gibson [30] and in Watanabe [31]. In Beck & Gibson [30], as well as in Gottheil & Bitterman [33], Epstein, Bontrager, & Park [34], and Wallach & Moore [35], the observers were shown triangles with a variety of shapes and responded by trying to construct a similar triangular shape. A comparison of the physical shapes of the triangles with the perceived shapes, as represented by their constructions, served as the measure of shape constancy. All of these studies showed that shape constancy was very poor during monocular viewing and that it only improved somewhat during binocular viewing. This discrepancy between the physical and the perceived shapes of the triangles with binocular viewing was also observed by Watanabe [31]. Watanabe claimed that this occurred because the perceived space was distorted. However, his failure to achieve perfect, or near perfect, shape constancy can be explained more parsimoniously by the geometrical ambiguity of the visual stimuli used and not by defects in the visual systems of the observers.

3. General Discussion

This study examined how the shape and orientation of a triangle within a 3D scene can be recovered from its 2D retinal image. The orientation of the triangle can be arbitrary unless the normal to the plane of the triangle is perpendicular to any of the lines of the

projection. Almost any triangular shape can be projected to the retinal image if the image is less than 20° (see Monte-Carlo simulation; $\theta_{\max} \leq 20^\circ$ in Figure 2A,B, Figures 3A and 4A). This ambiguity cannot be resolved even when the triangle is viewed binocularly (see our analyses of the images of the triangles that served as our visual stimuli; Figure 4B). When the retinal image is large, some shapes cannot be projected to the retinal image (see Monte-Carlo simulation; $\theta_{\max} > 20^\circ$ in Figure 2A,B and Figure 3B). This suggests that the retinal image of a triangle cannot serve as a cue for the shape of the triangle unless it is sufficiently large.

Now consider that if there are 4 feature points in a 3D scene, they usually form a volumetric polyhedron that has 4 vertices. This brings up the Perspective-4-Point (P4P) problem where one must recover the orientation of the polyhedron when the 3D shape and the 2D retinal image of the polyhedron are given [36]. It has been proven that there are 0 or up to 5 possible 3D interpretations of the polyhedron for a given shape, as well as for the size, and for the retinal image of the given polyhedron. Now, consider a case in which the 4 points are constrained to be coplanar to one another in the scene, and in which they form a planar polygon. Their orientation can be uniquely recovered from its monocular retinal image when the shape of the polygon is given [37,38]. A recovery is also usually possible from a stereo-pair of retinal images under the constraint that the shape is planar, but the shape of the polygon need not be given [39].

There needs to be 5, or more than 5, feature points in a 3D scene that project to a stereo-pair of retinal images if the 3D scene is going to be recovered from the stereo-pair [40,41]. A triangle has only 3 vertices so these are not sufficient to execute a recovery. The recovery becomes possible if there are 2, or more than 2, additional feature points in the scene [42]. A shortage of visual information on the retinas can also be compensated by using oculomotor information about the orientations of the eyes relative to the head. Note that the orientations of the eyes can be estimated by using the efference signal produced by the oculomotor control system.

Other “simple” 3D visual stimuli, such as an ellipse [43], and points on the sagittal plane that bisects the interocular axis perpendicularly [39], as well as points on a plane coplanar with the eyes [44], present analogous problems. Note that many psychophysical studies have shown that perception is not veridical and that percepts are distorted when such simple visual stimuli are used (e.g., [32]). However, note that such distortions could be attributed to defects in another mechanism that is being used to compensate for the shortage of visual information inherent in the too simple visual stimuli, such as the oculomotor efference signal. Note that other studies have shown that our perception of 3D scenes and the shapes of 3D objects is veridical in everyday life [4,45].

In our everyday life, 3D scenes “out there” are complex and it is this complexity that plays the critical role in perceiving them veridically. There are usually many feature points in a natural 3D scene and these points are essential for the binocular recovery of 3D [45,46] particularly when they become more widely distributed [44,47,48] and if they have sufficient density [49]. These feature points are inherent in the 3D objects present in the scene. Note that: (i) the shapes and positions of these objects often satisfy a number of a priori constraints, and (ii) the visual system can make use of these a priori constraints to recover a 3D scene from its 2D representation on the retina [4–7] (see Reference [50] for example). In light of these facts, it is questionable that many, if any, inferences about the visual system can be generalized to the veridical perception observed in our everyday life from the non-veridical perceptions that have been observed in many psychophysical studies that used very simple visual stimuli, such as triangles, ellipses, and planes that intersect the interocular axis perpendicularly, or are coplanar with the eyes. The visual information required to recover 3D scenes is absent when these simple visual stimuli are used. Human performance observed under such deprived conditions cannot be generalized to performance under natural viewing conditions.

These simple visual stimuli were selected and used because they provided a convenient way to eliminate artifacts from the visual stimuli and to facilitate control of the

experiment. The goal of our study was to *explain why* using a triangle introduces ambiguity that does not exist when more complex stimuli are used. The results of our simulation experiments show clearly that we achieved our goal. This encourages us to conclude by emphasizing that understanding the theoretical properties of one's visual stimuli is critical for designing experiments concerned with shape and depth and for interpreting the results obtained.

Author Contributions: Conceptualization, T.S.; methodology, V.M. and T.S.; software, V.M. and T.S.; validation, V.M. and T.S.; formal analysis, T.S.; investigation, V.M. and T.S.; resources, T.S.; data curation, V.M. and T.S.; writing—original draft preparation, V.M. and T.S.; writing—review and editing, T.S.; visualization, V.M. and T.S.; supervision, T.S.; project administration, T.S.; funding acquisition, T.S. All authors have read and agreed to the published version of the manuscript.

Funding: The article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE University) in 2019 (grant № 19-04-006) and by the Russian Academic Excellence Project «5–100».

Data Availability Statement: The C++ code used in this study is openly available in GitHub [51].

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The Perspective-3-Point (P3P) problem is concerned with recovering the orientation of a triangle ABC from its retinal image when the 3D shape of ABC and its 2D retinal image are given [19–28]. There are 0 or up to 4 possible 3D interpretations of the triangle. An algorithm developed by Fischler & Bolles [21] for solving the P3P problem was implemented in a C++ program and used for the simulations in this study. This algorithm recovers the distance of the vertices A , B , and C of the triangle from a center of projection E . Let the distance of the vertices A , B , and C be l_A , l_B , and l_C . The output of this algorithm is vectors with 3 values. These vectors represent individual 3D interpretations of the triangle and the 3 values of the vectors represent l_A , l_B , and l_C .

There are two important properties of the algorithm that were used in our implementation. First, the algorithm needs to solve a quartic equation for recovering l_A , l_B , and l_C . The roots of this quartic equation were computed by using Ferrari's method [52]. Ferrari's method was implemented in a C++ function based on [53].

Second, the algorithm recovers the distance l_A , l_B , and l_C of the vertices A , B , and C in an unbalanced manner. The distance to 2 of the 3 vertices is recovered first and the distance to the last vertex is recovered on the basis of the recovered distance to the first 2 vertices. In a few rare cases, the results of the recovery changes depending on which vertex was recovered last because of rounding and discretization errors. This problem was addressed in our implementation by recovering l_A , l_B , and l_C with 3 different orders in which the last recovered distance was different from one another. Note that a result of the recovery was vectors of the recovered l_A , l_B , and l_C . Sets of the recovered vectors from the 3 different orders were combined and duplications of the vectors were eliminated. The vectors were also verified by recovering the 3D shape of ABC from its 2D retinal image and from each recovered vector. The vector was eliminated whenever the recovered shape was substantially different from the given shape of ABC . The C++ code for this implementation was uploaded to GitHub [51].

Our implementation of the algorithm was tested in a simulation experiment. In each trial of this experiment, a triangle was randomly generated in a 3D scene and its retinal image was computed. The 3D XYZ Cartesian coordinate system was set in the 3D scene to place the origin at the center of projection E . The vertices A , B , and C were placed in the scene so that their Z -coordinates were between 1 and 100. Angles between the Z -axis and lines of projection to A , B , and C were less than ε , where ε is 45° in one condition and is 85° in the other condition, and 10^6 scenes were randomly generated for each condition.

In each trial, the algorithm recovered the depth l_A , l_B , and l_C of the triangle's vertices A , B , and C from the retinal image of the simulated scene and the simulated shape of the

triangle. The recovered depth was compared with the simulated depth of the vertices as follows:

$$\delta = \left(\frac{l_A}{\|l\|} - \frac{l'_A}{\|l'\|} \right)^2 + \left(\frac{l_B}{\|l\|} - \frac{l'_B}{\|l'\|} \right)^2 + \left(\frac{l_C}{\|l\|} - \frac{l'_C}{\|l'\|} \right)^2 \begin{cases} \|l\| = \sqrt{l_A^2 + l_B^2 + l_C^2} \\ \|l'\| = \sqrt{l'^2_A + l'^2_B + l'^2_C} \end{cases} \quad (A1)$$

where $l_A, l_B,$ and l_C are the simulated depth of the vertices $A, B,$ and C and $l'_A, l'_B,$ and l'_C are their recovered depth. The recovered depth was considered to be veridical if the difference δ between the simulated and recovered depth was less than 0.01.

The depth of the same 10^6 scenes were also recovered by using 3 other algorithms that are used to solve the P3P problem, specifically, Gao, Hou, Tang, & Chang [22]; Ke & Roumeliotis [25]; and Banno [27]. The algorithms of Gao, Hou, Tang, & Chang [22] and Ke & Roumeliotis [25] were implemented as a function in the OpenCV library (ver. 4.2.0, [54,55]). The algorithm of Banno [27] was implemented by Banno himself [27]. The performance of all of these 3 implementations was compared with our implementation.

The performance of these implementations is shown in Table A1. Their performance was evaluated by examining the probability of failing to recover the depth of the simulated scenes and the time required to recover the scenes. Note that the retinal images were projections of the simulated scenes and the simulated shapes of the triangles were given for the recovery. Hence, possible interpretations of the retinal images should have included the simulated scenes. Our implementation always recovered the simulated 3D scene from its retinal image. The other 3 algorithms failed to recover some of the simulated scenes. Our implementation’s processing speed was also the highest.

Table A1. Performance of the 4 implementations of the P3P algorithms used to recover a 3D scene with a randomly-generated triangle. The implementations were evaluated on the basis of their recovery of 10^6 scenes. The probability of the failure to recover the simulated scene and the time required for processing the recovery of the 10^6 scenes were measured.

Algorithms	ϵ	% Failed Trial	Processing Time
Minkov & Sawada	45°	0%	2.7 s
	85°	0%	2.6 s
Gao, Hou, Tang, & Chang [22]	45°	0.034%	20 s
	85°	0.025%	18 s
Ke & Roumeliotis [25]	45°	21%	25 s
	85°	25%	26 s
Banno [27]	45°	29%	8.7 s
	85°	22%	7.8 s

The performance of our implementation was also tested with different values of the threshold for the depth difference δ . The implementation continued to perform perfectly when the threshold was set to 10^{-6} or larger and only a few errors (0.0041%) were made when the threshold was set to 10^{-7} . Our implantation of the algorithm used for solving the P3P problem was highly reliable.

References

1. Watt, R.J. Towards a general theory of the visual acuities for shape and spatial arrangement. *Vis. Res.* **1984**, *24*, 1377–1386. [CrossRef]
2. Campbell, F.W.; Robson, J.G. Application of Fourier analysis to the visibility of gratings. *J. Physiol.* **1968**, *197*, 551–566. [CrossRef]
3. Pizlo, Z. *3D Shape: Its Unique Place in Visual Perception*; MIT Press: Cambridge, MA, USA, 2008.
4. Pizlo, Z.; Li, Y.; Sawada, T.; Steinman, R.M. *Making a Machine That Sees Like Us*; Oxford University Press: New York, NY, USA, 2014.
5. Pizlo, Z.; Sawada, T.; Li, Y.; Kropatsch, W.; Steinman, R.M. New approach to the perception of 3D shape based on veridicality. Complexity, Symmetry and Volume. *Vis. Res.* **2010**, *50*, 1–11. [CrossRef]

6. Sawada, T.; Li, Y.; Pizlo, Z. Shape perception. In *Oxford Handbook of Computational and Mathematical Psychology*; Busemeyer, J., Townsend, J., Wang, Z.J., Eidels, A., Eds.; Oxford University Press: New York, NY, USA, 2015; pp. 255–276.
7. Poggio, T.; Torre, V.; Koch, C. Computational vision and regularization theory. *Nature* **1985**, *317*, 314–319. [[CrossRef](#)]
8. Howard, I.P. Perceiving in Depth. In *Other Mechanisms of Depth Perception*; Oxford University Press: Oxford, UK, 2012.
9. Mishkin, M.; Ungerleider, L.G.; Macko, K.A. Object vision and spatial vision: Two cortical pathways. *Trends Neurosci.* **1983**, *6*, 414–417. [[CrossRef](#)]
10. Kravitz, D.J.; Saleem, K.S.; Baker, C.I.; Mishkin, M. A new neural framework for visuospatial processing. *Nat. Rev. Neurosci.* **2011**, *12*, 217–230. [[CrossRef](#)] [[PubMed](#)]
11. Scharff, A.; Palmer, J.; Moore, C.M. Divided attention limits perception of 3-D object shapes. *J. Vis.* **2013**, *13*, 18. [[CrossRef](#)]
12. Pentland, A.P. Perceptual organization and the representation of natural form. *Artif. Intell.* **1986**, *28*, 293–331. [[CrossRef](#)]
13. Marr, D. *Vision*; W.H. Freeman: New York, NY, USA, 1982.
14. Biederman, I. Recognition-by-components: A theory of human image understanding. *Psychol. Rev.* **1987**, *94*, 115–147. [[CrossRef](#)]
15. Leeuwenberg, E.; van der Helm, P.A. *Structural Information Theory: The Simplicity of Visual Form*; Cambridge University Press: New York, NY, USA, 2013.
16. Bogen, J. Theory and Observation in Science. In *The Stanford Encyclopedia of Philosophy*, Winter 2020 ed.; Zalta, E.N., Ed.; Stanford University: Stanford, CA, USA, 2020; Available online: <https://plato.stanford.edu/archives/win2020/entries/science-theory-observation/> (accessed on 28 December 2020).
17. Brewer, W.F.; Lambert, B.L. The theory-ladenness of observation and the theory-ladenness of the rest of the scientific process. *Philos. Sci.* **2000**, *68*, S176–S186. [[CrossRef](#)]
18. Runeson, S. On the possibility of “smart” perceptual mechanisms. *Scand. J. Psychol.* **1977**, *18*, 172–179. [[CrossRef](#)]
19. Lu, X.X. A review of solutions for perspective-n-point problem in camera pose estimation. *J. Phys. Conf. Ser.* **2018**, *1087*, 052009. [[CrossRef](#)]
20. Haralick, B.M.; Lee, C.N.; Ottenberg, K.; Nölle, M. Review and analysis of solutions of the three point perspective pose estimation problem. *Int. J. Comput. Vis.* **1994**, *13*, 331–356. [[CrossRef](#)]
21. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
22. Gao, X.; Hou, X.; Tang, J.; Cheng, H. Complete solution classification for the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 930–943.
23. Li, S.; Xu, C. A stable direct solution of perspective-three-point problem. *Int. J. Pattern Recognit. Artif. Intell.* **2011**, *25*, 627–642. [[CrossRef](#)]
24. Grafarend, E.W.; Shan, J. Closed-form solution of P4P or the three-dimensional resection problem in terms of Möbius barycentric coordinates. *J. Geod.* **1997**, *71*, 217–231. [[CrossRef](#)]
25. Ke, T.; Roumeliotis, S.I. An efficient algebraic solution to the perspective-three-point problem. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4618–4626. [[CrossRef](#)]
26. Nakano, G. *A Simple Direct Solution to the Perspective-Three-Point Problem*, Proceedings of the 30th British Machine Vision Conference (BMVC 2019), Cardiff, UK, 9–12 September 2019; BMVA Press: Durham, UK, 2019; Volume 26, pp. 1–12.
27. Banno, A. A P3P problem solver representing all parameters as a linear combination. *Image Vis. Comput.* **2018**, *70*, 55–62. [[CrossRef](#)]
28. Persson, M.; Nordberg, K. Lambda twist: An accurate fast robust perspective three point (P3P) solver. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 318–332.
29. DeMenthon, D.; Davis, L.S. Exact and approximate solutions of the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 1100–1105. [[CrossRef](#)]
30. Beck, J.; Gibson, J.J. The relation of apparent shape to apparent slant in the perception of objects. *J. Exp. Psychol.* **1955**, *50*, 125–133. [[CrossRef](#)]
31. Watanabe, T. The estimation of the curvature of visual space with a visual triangle. *Jpn. J. Psychol.* **1996**, *67*, 278–284. [[CrossRef](#)] [[PubMed](#)]
32. Indow, T. *The Global Structure of Visual Space*; World Scientific: River Edge, NJ, USA, 2004; Volume 1.
33. Gottheil, E.; Bitterman, M.E. The measurement of shape-constancy. *Am. J. Psychol.* **1951**, *64*, 406–408. [[CrossRef](#)] [[PubMed](#)]
34. Epstein, W.; Bontrager, H.; Park, J. The induction of nonveridical slant and the perception of shape. *J. Exp. Psychol.* **1962**, *63*, 472–479. [[CrossRef](#)]
35. Wallach, H.; Moore, M.E. The role of slant in the perception of shape. *Am. J. Psychol.* **1962**, *75*, 289–293. [[CrossRef](#)]
36. Gao, X.S.; Tang, J. On the probability of the number of solutions for the P4P problem. *J. Math. Imag. Vis.* **2006**, *25*, 79–86. [[CrossRef](#)]
37. Abidi, M.A.; Chandra, T. A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1995**, *17*, 534–538. [[CrossRef](#)]
38. Pizlo, Z. A theory of shape constancy based on perspective invariants. *Vis. Res.* **1994**, *34*, 1637–1658. [[CrossRef](#)]
39. Longuet-Higgins, H.C. The reconstruction of a plane surface from two perspective projections. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* **1986**, *227*, 399–410.

40. Kruppa, E. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung [To determine a 3D object from two perspective views with known inner orientation]. *Sitz. Der Math. Nat. Kais. Akad. Wiss.* **1939**, *122*, 1939–1948, (Translated by Gallego, G., Mueggler, E., Sturm, P. arXiv **2017**, <https://arxiv.org/abs/1801.01454>).
41. Thompson, E.H. A rational algebraic formulation of the problem of relative orientation. *Photogramm. Rec.* **1959**, *3*, 152–159. [[CrossRef](#)]
42. Sawada, T. A Computational Model that recovers depth from stereo-input without using any oculomotor information. *J. Math. Psychol.* under review. [[CrossRef](#)]
43. Pizlo, Z.; Salach-Golyska, M. Is vision metric? Comment on Lappin and Love (1992). *Percept. Psychophys.* **1994**, *55*, 230–234. [[CrossRef](#)]
44. Backus, B.; Banks, M.S.; van Ee, R.; Crowell, J.A. Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vis. Res.* **1999**, *39*, 1143–1170. [[CrossRef](#)]
45. Kwon, T.; Li, Y.; Sawada, T.; Pizlo, Z. Gestalt-like constraints produce veridical (Euclidean) percepts of 3D indoor scenes. *Vis. Res.* **2016**, *126*, 264–277. [[CrossRef](#)] [[PubMed](#)]
46. Peek, S.A.; Mayhew, J.E.; Frisby, J.P. Obtaining viewing distance and angle of gaze from vertical disparity using a Hough-type accumulator. *Image Vis. Comput.* **1984**, *2*, 180–190. [[CrossRef](#)]
47. Kaneko, H.; Howard, I.P. Spatial limitation of vertical-size disparity processing. *Vis. Res.* **1997**, *37*, 2871–2878. [[CrossRef](#)]
48. Bradshaw, M.F.; Glennerster, A.; Rogers, B.J. The effect of display size on disparity scaling from differential perspective and vergence cues. *Vis. Res.* **1996**, *36*, 1255–1264. [[CrossRef](#)]
49. Gantz, L.; Bedell, H.E. Variation of stereothreshold with random-dot stereogram density. *Optom. Vis. Sci.* **2011**, *88*, 1066–1071. [[CrossRef](#)] [[PubMed](#)]
50. Erkelens, C.J. Evidence for obliqueness of angles as a cue to planar surface slant found in extremely simple symmetrical shapes. *Symmetry* **2015**, *7*, 241–254. [[CrossRef](#)]
51. Sawada, T. P3P. 2020. Available online: <https://github.com/TadamasaSawada/P3P> (accessed on 22 January 2020).
52. Weisstein, E.W. Quartic Equation. 2020. Available online: <http://mathworld.wolfram.com/QuarticEquation.html> (accessed on 13 February 2021).
53. Khasin, S.I. Решение уравнения 3-й, 4-й и 5-й степеней на C++ [Solution of the 3rd, 4th, and 5th degree polynomial equations in C ++]. 2018. Available online: <http://math.ivanovo.ac.ru/dalgebra/Khashin/cutil/poly34.html> (accessed on 16 June 2018).
54. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* **2000**.
55. Itseez. Open Source Computer Vision Library. 2015. Available online: <https://github.com/itseez/opencv> (accessed on 7 January 2020).