NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Vasily Minkov, Tadamasa Sawada

SEEING A TRIANGLE IN A 3D SCENE MONOCULARLY AND BINOCULARLY

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: PSYCHOLOGY
WP BRP 91/PSY/2018

Vasily Minkov[1], Tadamasa Sawada[2]

SEEING A TRIANGLE IN A 3D SCENE MONOCULARLY AND BINOCULARLY[3]

Theoretical understanding of a visual stimulus in a psychophysical experiment is critical for controlling the experiment and for interpreting its results. This fact encourages vision scientists to use "simple" visual stimuli. A triangle in a 3D scene is one of the simplest stimuli for studying 3D perception. In this study, we analyzed geometrical properties of a relation between the triangle and its retinal image using a computer algorithm. Based on the analysis, we discuss validity of results of past studies that used triangles as their visual stimuli.

Keywords: shape constancy; shape ambiguity; visual space; Euclidean geometry; non-Euclidean geometry; binocular disparity; P3P problem.

JEL Classification: Z

PsycINFO Classification Categories and Codes: 2323 Visual Perception

[1] Undergraduate Student, proveyourselfmail@gmail.com, School of Psychology, National Research University Higher School of Economics.
[2] Assistant Professor, PhD, tsawada@hse.ru (tada.masa.sawada@gmail.com), School of Psychology, National Research University Higher School of Economics.
[3] This manuscript was prepared as a part of a coursework project of the first author.
This Working Paper is an output of a research project implemented within NRU HSE's Annual Thematic Plan for basic and applied research. Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

Introduction

Theoretical understanding of a visual stimulus and of an observer's task in a psychophysical experiment is critical for controlling the experiment and for interpreting results of the experiment. This fact encourages vision scientists to use "simple" visual stimuli because the simple visual stimuli are expected to make understanding the results easier. Using simple visual stimuli can be also justified by "reductionism". A visual proximal stimulus (a 2D retinal image) is often decomposed into points, contours, gratings, and Gabor patterns and perception of them has been studied in psychophysics (e.g. Watt, 1984; Campbell & Robson 1968).

Perception of 3D spatial properties of visual distal stimuli (3D scenes and 3D shapes) cannot be studied in the same way as perception of 2D spatial properties of proximal visual stimuli (Pizlo, 2008). The visual system is given 2D projections of the 3D distal stimuli but is not the distal stimuli themselves. Projecting a 3D distal stimulus to a 2D retina is a well-posed forward problem. The retinal image can be computed from the distal stimulus with a viewing condition (like 3D computer graphics). On the other hand, it is an ill-posed inverse problem to recover the 3D distal stimulus from the 2D retinal image. There are infinitely many possible 3D interpretations of the retinal image. The visual system can resolve this problem in two different ways: using a priori constraints about the 3D distal stimulus (Poggio, Torre, & Koch, 1985; Pizlo, Li, Sawada, & Steinman, 2014) and using depth cues (Howard, 2012).

Visual distal stimuli (3D scenes and 3D shapes) are often represented as compositions of points or of triangles in computer vision and computer graphics. The triangle has been used for representing the scenes and the shapes because it is the simplest polygon with a surface and is always planar. Then, understanding properties of the triangle and of its 2D projection is important for discussing results of psychophysical studies using triangles as their visual stimuli and for studying perception of more complex 3D structures composed of triangles.

In this study, we will numerically analyze i) geometrical properties of a relation between a triangle in a 3D scene and its retinal image and ii) retinal images of triangles used as visual stimuli in past psychophysical studies using an existing algorithm of geometry. Based on the analysis, we will discuss validity of results in the past psychophysical studies.

Analysis

A relation between a triangle *ABC* in a 3D scene and its 2D perspective projection (see Appendix for a 2D orthographic projection of *ABC*) to a retina can be represented by a tetrahedron *EABC* (Figure 1). A bottom face of the tetrahedron is the triangle *ABC* and an apex *E* represents a center of projection of an eye. A retinal image of *ABC* can be represented by three visual angles θ_{BC}, θ_{CA}, θ_{AB} at *E*. A 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 $\omega_C = 180° - \omega_A - \omega_B$. A size of *ABC* can be controlled by a length of the line segment *AB*. The length of *AB* can be fixed to be 1 without loss of any generality. If the size of *ABC* is larger by a factor of *s*, the size of the tetrahedron *EABC* becomes also larger by a factor of *s* while all the angles of *EABC* remain constant.

Consider a shape of the triangle *ABC* and its retinal image are given and recovering a position and orientation of the triangle from the given information. This problem has been referred as Perspective-3-Point (P3P) problem (e.g. Fischler & Bolles, 1981; Gao, Hou, Tang, & Cheng, 2003; Li & Xu, 2011). It has been proved that there are up to 4 possible 3D interpretations of the triangle for the given shape of and the given retinal image of the triangle.

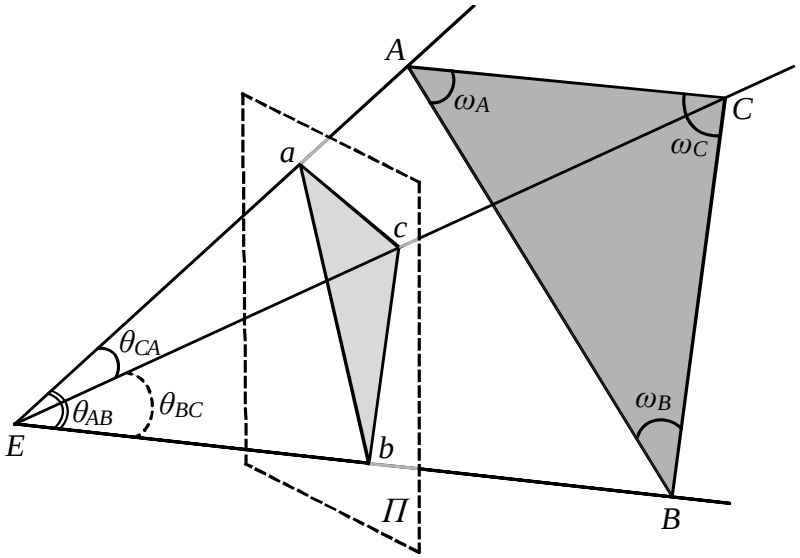

Figure 1. A perspective projection from a triangle *ABC* in a 3D scene to a triangle *abc* on a 2D image plane *Π* from a center of projection *E*. The projection can be represented as a tetrahedron *EABC*.

Monte-Carlo simulation

We tested frequencies of numbers of possible 3D interpretations of a triangle for a retinal image in two experiments of Monte-Carlo simulation. A shape of the triangle (ω_A, ω_B, ω_C) and the retinal image (θ_{BC}, θ_{CA}, θ_{AB}) were randomly generated in each trial by randomly sampling ω_A, ω_B, θ_{BC}, θ_{CA}, and θ_{AB} from uniform distributions. The sampled variables of ω_A, ω_B, θ_{BC}, θ_{CA}, and θ_{AB} were constrained so that ω_A, ω_B, and ω_C form the triangle and θ_{BC}, θ_{CA}, and θ_{AB} form an apex of the tetrahedron: $\omega_A + \omega_B + \omega_C = 180°$, $\theta_{BC} + \theta_{CA} + \theta_{AB} < 360°$, $\theta_{BC} + \theta_{CA} > \theta_{AB}$, $\theta_{CA} + \theta_{AB} < \theta_{BC}$, $\theta_{AB} + \theta_{BC} < \theta_{CA}$. Additionally, a shape of the triangle was restricted by another constraint: $10° < \omega_A, \omega_B, \omega_C < 170°$. Then, possible 3D interpretations of the triangle for the retinal image (θ_{BC}, θ_{CA}, θ_{AB}) are computed using an algorithm of Fischler and Bolles (1981).

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

Results of the simulation are shown in Figure 2. The ordinates show frequencies of numbers of possible 3D interpretations. The abscissa shows $\theta_{\max}$, which controls the range of the sampling. The four curves indicate the numbers of possible 3D interpretations.

The results show that the frequency of 2 possible interpretations is almost 100% (> 95%) if the visual angles θ_{BC}, θ_{CA}, and θ_{AB} are small ($\theta_{\max} \leq 14°$ in Figure 2AB) and decreases as the retinal image becomes larger. The number of possible interpretations is mostly 0 or 1 (> 94%) if all the visual angles θ_{BC}, θ_{CA}, and θ_{AB} are larger than 35° ($\theta_{\max} \geq 70°$ in Figure 2B). The number is rarely 3 or 4 for any value of $\theta_{\max}$.

Note that a perspective projection within a region of a small visual angle around the fovea can be well approximated by an orthographic projection with uniform scaling (see Appendix). Number of possible 3D interpretations of the triangle *ABC* for a given orthographic projection is always 2. It explains why the number of possible interpretations is almost always 2 under the perspective projection when the visual angles θ_{BC}, θ_{CA}, and θ_{AB} are small. As the visual angles become larger, discrepancy between the perspective and the orthographic projections becomes larger.

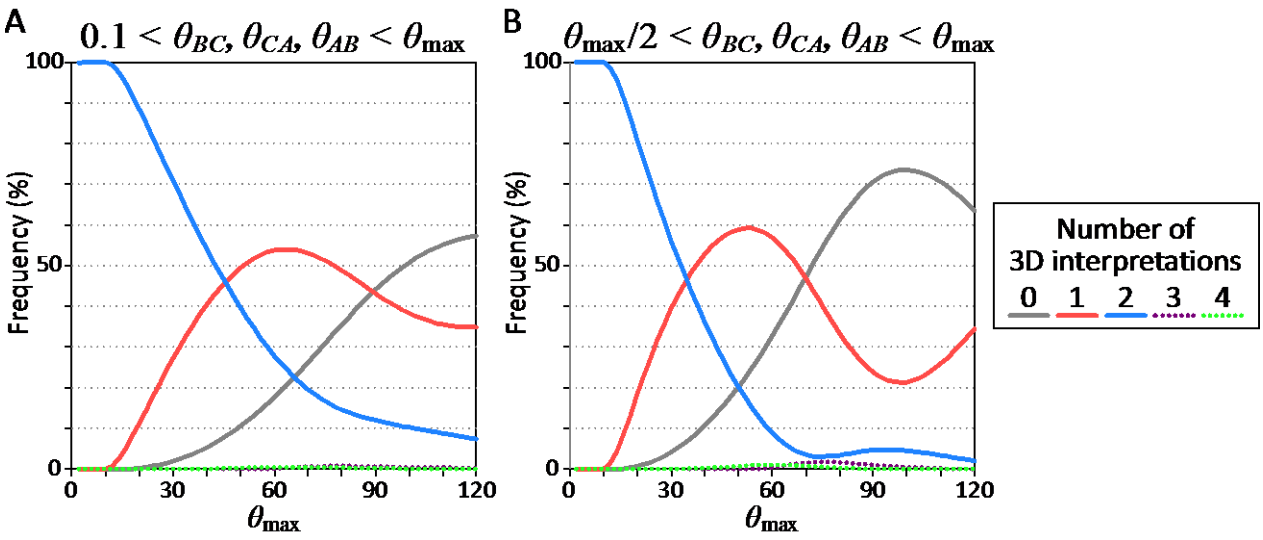

Figure 2. Results from the two experiments of Monte-Carlo simulation. The ordinate shows frequency of numbers of possible 3D interpretations and the abscissa shows θ_{max}. The five curves indicate the numbers of possible 3D interpretations. (A) The visual angles θ_{BC}, θ_{CA}, and θ_{AB} were sampled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

Past visual stimuli of triangles

Some psychophysical studies used triangles as their visual stimuli for testing human performance in shape constancy (Gottheil & Bitterman, 1951; Epstein, Bontrager, & Park, 1962; Wallach & Moore, 1962; Beck & Gibson, 1955; see also Pizlo, 1994, 2008) and for measuring discrepancy of the visual space from Euclidean space (Watanabe, 1996, see also Indow, 2004). These studies compared a shape of a triangle shown to an observer and a perceived shape of the triangle by the observer. Difference between these shapes was interpreted as failure of shape constancy in the shape constancy studies and was interpreted as an evidence of non-Euclidean visual space in Watanabe (1996).

We analyzed the retinal images of the triangles in Experiment 1 of Beck and Gibson (1955) and in Condition 3 of Watanabe (1996) using the algorithm of P3P problem and shapes of triangles that can be projected to these retinal images were computed. Note that these studies were chosen because of simplicity of the stimuli and clarity of their apparatus settings. Their experiments were conducted in darkrooms and the triangles were shown without any other visual information. The triangles were observed monocularly in Beck and Gibson (1955) and binocularly in Watanabe (1996).

Results of the analysis on the retinal images are shown in Figure 3. Each point in these maps represents a shape of a triangle. The abscissas and the ordinates show the two angles of the triangle ω_A and ω_B. The third angle ω_C of the triangle can be computed as ω_C = 180° - ω_A - ω_B. Colors indicate number of possible interpretations of the triangle with some shape (ω_A, ω_B, ω_C) for the retinal image. The retinal image of the triangle under monocular observation is always ambiguous (Figure 3A). The triangle with most of shapes can be projected to the same retinal image.

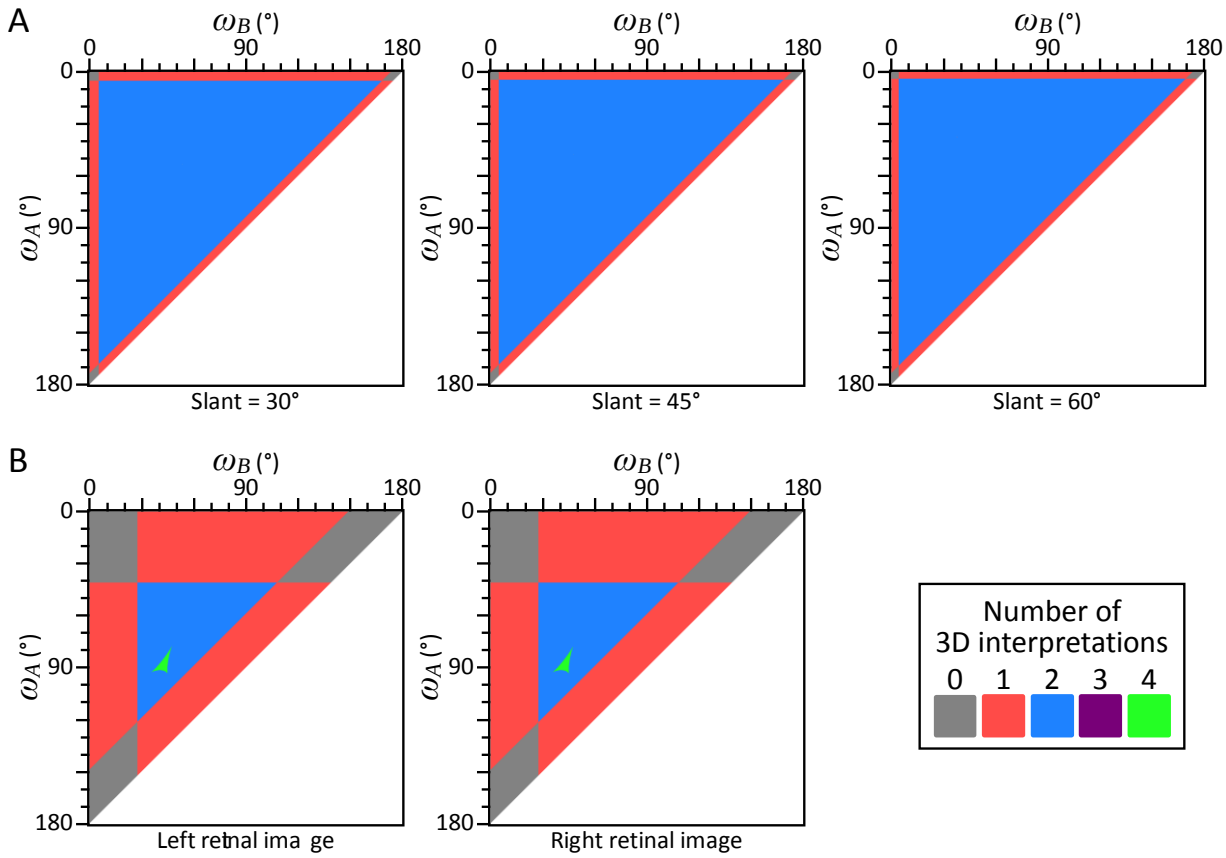

Figure 3. Results of analysis on the retinal images of triangles in (A) Beck and Gibson (1955) and (B) Watanabe (1996). The ordinate and abscissa show ω_A and ω_B. Colors indicate number of possible interpretations of the triangle with a shape (ω_A, ω_B, 180° − ω_A − ω_B) for the retinal image. White regions represent invalid triangles in Euclidean space: ω_A + ω_B ≥ 180°. (A) The analysis results on the retinal images in Experiment 1 of Beck and Gibson (1955) under 3 conditions of forward slant (30, 45, and 60°) of the triangle. The shape of the triangle (ω_A, ω_B, ω_C) is (57.995°, 57.995°, 64.010°). The retinal images (θ_{BC}, θ_{CA}, θ_{AB}) are (5.538°, 5.538°, 6.573°) for 30°, (4.928°, 4.928°, 6.638°) for 45°, and (4.222°, 4.222°, 6.689°) for 60° of the slant. (B) The analysis results on the retinal images in Condition 3 of Watanabe (1996) under binocular observation. The shape of the

triangle (ω_A, ω_B, ω_C) is (87.459°, 42.946°, 49.594°). The retinal images (θ_{BC}, θ_{CA}, θ_{AB}) are (41.19°, 27.68°, 31.01°) for a left eye and (41.18°, 27.78°, 30.91°) for a right eye.

This ambiguity cannot be resolved even under binocular observation (Figure 3B). The left and right panels of Figure 3B show ambiguity of the left and right retinal images of a single triangle in Condition 3 of Watanabe (1996). These two panels are almost identical with one another. A stereo pair of the retinal images are individually ambiguous and the triangle with almost any shape that can be projected to one of the retinal images can be projected to the other retinal image as well.

The studies of shape constancy using a triangle visual stimulus found that a perceived shape of the triangle was very different from its veridical shape unless some depth cue (e.g. binocular disparity) was provided. It is reasonable because a monocular retinal image of the triangle is consistent with many different shapes of triangles (Figure 3A). They concluded that depth information is critical for shape constancy. However, human performance in shape constancy is actually reliable even under monocular condition if a shape is not a triangle but is complex and regular enough (Pizlo, 1994; Li & Pizlo, 2011; Liu, Knill, & Kersten, 1995). The poor performance in shape constancy of a triangle can be explained by its geometrical property. However, Figure 3B shows that the shape of the triangle is still theoretically ambiguous even under binocular observation.

This ambiguity can be resolved if the visual system knows how a stereo pair of eyes are oriented relative to their interocular axis. It is theoretically possible to estimate the eye orientations only from their retinal images if there are 5 or more than 5 feature points in the scene that are projected to the both images (Kruppa, 1913, translated by Gallego, Meggler, & Sturm, 2017; Thompson, 1959; see Hartley & Zisserman, 2003 for a review). However, the triangle has only 3 vertices and they are not enough for estimating the eye orientations.

The visual system can compensate for this shortage of visual information by using the efference signal from the oculomotor control system (Foley, 1980; Bradshaw, Glennerster, & Rogers, 1996; Backus, Banks, van Ee, & Crowell, 1999). The oculomotor efference signal can tell how the two eyes are oriented. Then, depth of even a single point in a 3D scene can be recovered with the oculomotor information (vergence as a depth cue, see Howard, 2012 for a review). However, the oculomotor efference signal can play any major role on binocular depth perception only if a visual stimulus is impoverished. It is unlikely in our everyday life because a 3D scene out

there is with many feature points and provides rich visual information in a wide visual field (see Discussion).

Triangle visual stimuli used in the shape constancy studies (Gottheil & Bitterman, 1951; Epstein, Bontrager, & Park, 1962; Wallach & Moore, 1962; Beck & Gibson, 1955) were all smaller than 10°.[4] These visual stimuli are too small for the visual system to recover their 3D information without the oculomotor efference signal (Backus et al., 1999; Bradshaw, Glennerster, & Rogers, 1996).

Triangle stimuli used in Watanabe (1996) were sufficiently large; the visual angles between the vertices of the triangles were around 31° on average (Figure 4). Additionally, several extra points were shown with the triangles depending on tasks of his experiment (6 extra points at maximum). Computationally, these points (both the extra points and the vertices of the triangles) provide sufficient information for recovering 3D information from the stimuli. However, it is questionable whether the human visual system can integrate such sparse points in the wide visual field to recover 3D information only from the retinal images (Kaneko & Howard, 1997; Zhang, Berends, & Schor, 2003; Gantz & Bedell, 2011). Then, the visual system should rely again on the oculomotor efference signal but the efference signal may introduce systematic error on visual perception (Mistudo, 2007; Mitsudo, Kaneko, & Nishida, 2009). Note that Watanabe (1996) reported distortion of 3D perception and concluded that the visual space is non-Euclidean. However, the distortion could be attributed to impoverishness of the visual stimuli.

[4] Epstein, Bontrager, and Park (1962) showed a textured background with the triangle and the background size was less than 16°. Beck and Gibson (1955) also showed a textured background with the triangle in their Experiments 2 and 3. However, the size of their background was not mentioned. On the other hand, Erkelens (2015) tested a perceived shape of a triangle in a normally-illuminated room so that other visual features in a wide visual field were available.

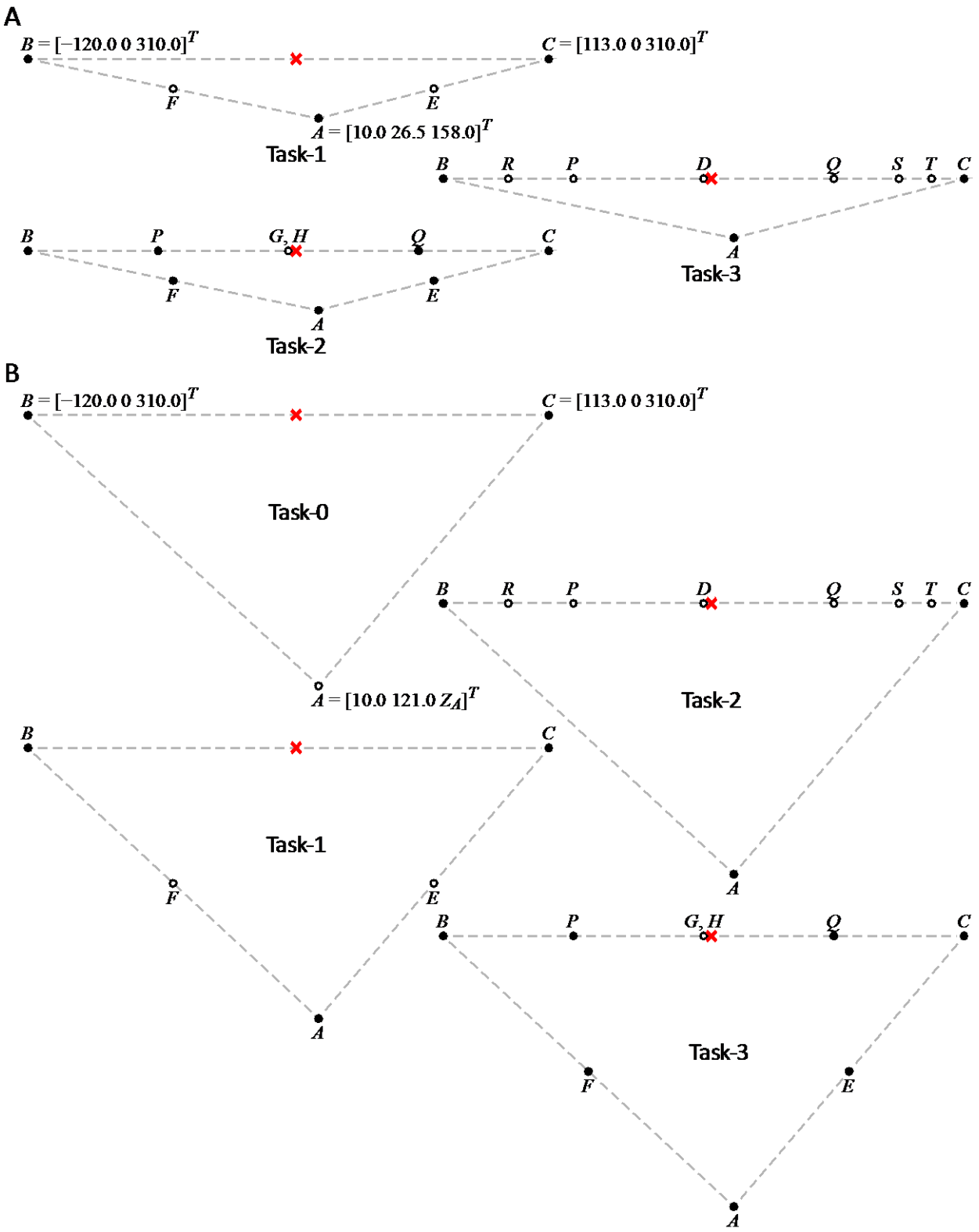

Figure 4. Orthographic projections of visual distal stimuli used in (A) Condition 2 and (B) Condition 3 of Watanabe (1996) along a direction of depth (the *Z*-axis). The *X*-axis is frontoparallel and horizontal (from left to right). The *Y*-axis is frontoparallel and vertical (its direction is not specified in Watanabe, 1996). A cyclopean eye of an observer was at $[0, 0, 0]^T$ and its orthographic projection

is indicated by x in the figure). The unit in the figure is centimeters. A triangle *ABC* is shown with extra points depending on tasks. Task-0 (only Condition 3): The *Z*-coordinate of *A* was adjusted so that the triangle *ABC* is frontoparallel. Task-1: the observer adjusted positions of *E* and *F* so that $E = (A+B)/2$ and $F = (A+C)/2$. Task-2: the observer adjusted positions of *D*, *P*, *Q*, *R*, *S*, and *T* so that $D = (B+C)/2$, $P = (B+D)/2$, $Q = (D+C)/2$, $R = (B+P)/2$, $S = (Q+C)/2$, and $T = (S+C)/2$. Task-3: the observer adjusted positions of *G* and *H* so that they are on a line segment *BC* and $|G-B| = |F-E|$ and $|H-C| = |F-E|$. Note that *G* and *H* were never shown together at the same time. Their geometrically correct positions in Euclid space coincide with one another.

Discussion

The computational analysis in this study showed that a triangle in a 3D scene is an "impoverished" visual stimulus even under binocular observation. There are many different 3D shapes of the triangle with many different 3D positions and orientations that can project to the retinal image and the stereo pair of retinal images.

It is difficult to discuss shape constancy (Gottheil & Bitterman, 1951; Epstein, Bontrager, & Park, 1962; Wallach & Moore, 1962; Beck & Gibson, 1955) and non-Euclidean visual space (Watanabe, 1996) based on results of their past studies using triangles as visual stimuli. The stimuli are so impoverished. Namely, they are too small, with few feature points, or with sparse feature points. In that case, the visual system should rely on the oculomotor efference signal to recover 3D information (see Past visual stimuli of triangles). However, the oculomotor efference signal should not play an important role for depth perception outside the laboratory because a 3D scene out there is full of visual information (see Backus et al., 1999; Bradshaw, Glennerster, & Rogers, 1996).

Recovering 3D information (3D shape or 3D scene) from a 2D retinal image is an ill-posed problem under monocular observation. However, our perception of a 3D scene out there is almost always veridical in our everyday life even under monocular observation. The everyday scene and objects there almost always satisfy some regularities (e.g. mirror-symmetry). Then, the visual system can recover their regular 3D information by applying *a priori* constraints of these regularities (Pizlo, Li, Sawada, & Steinman, 2014). The *a priori* constraints strongly bias our perception even under binocular observation (Li, Sawada, Shi, Kwon, & Pizlo, 2011; Jayadevan, Sawada, Delp, & Pizlo, accepted; Li & Pizlo, 2011; Sugihara, 2016). Erkelens (2015) reported a systematic bias of perception of a shape and of a 3D orientation of an isosceles triangle in a 3D scene. Such a bias toward a right angle is observed with some other objects (Perkins, 1972; Perkins, 1976; Sugihara, 1997, 2005, 2014a, b; Griffiths & Zaidi, 2000). It is worth pointing out that almost all triangles used

in the studies of shape constancy (Gottheil & Bitterman, 1951; Wallach & Moore, 1962; Beck & Gibson, 1955) were isosceles triangles.[5] Results reported in these studies could be affected by the right angle bias.

This study discussed specifically a triangle in a 3D scene because of uniqueness of its geometrical properties (e.g. the P3P problem). However, other "simple" 3D visual stimuli can have the same or analogous problems as the triangle does (e.g. Pizlo & Salach-Golyska, 1994). Perception of 3D spatial properties cannot be studied in the same way as perception of 2D spatial properties. The 3D properties are about a distal stimulus and the 2D properties are about a retinal image (a proximal stimulus), which is a projection of the distal stimulus. The visual system can perceive the 3D properties from the 2D retinal image but this is an ill-posed problem. Understanding this geometry is important for designing experiments for testing 3D perception and interpreting their results.

Appendix

A retinal image of a 3D scene is a 2D perspective projection. The perspective projection can be approximated within a region of a small visual angle around the fovea by a 2D orthographic projection along a line of sight with uniform scaling. In this appendix, a relation between a triangle in a 3D scene and its 2D orthographic projection is analyzed. Note that a general solution of this problem has been already shown in earlier studies (e.g. Hu & Ahuja, 1991; Pei & Liou, 1994). In this appendix, the problem is specifically formulated so that a 3D orientation of a triangle is represented by slant σ_{Δ}, tilt τ_{Δ}, and roll ρ_{Δ}. This representation is common in Psychophysics and makes the solution easier.

The *XYZ* Cartesian coordinate system of a 3D scene and the *xy* Cartesian coordinate system of a 2D image in the scene are set as follows: (i) the *Z*-axis of the 3D coordinate system is perpendicular to the image plane Π_I and Π_I is $Z = f$ where f is a constant and can be arbitrary, (ii) the *Z*-axis is normal to Π_I and intersect with Π_I at the origin of the 2D coordinate system, and (iii) the *X*- and *Y*-axes of the 3D coordinate system are parallel to the *x*- and *y*-axes of the 2D coordinate system, respectively. A 2D orthographic projection of a point $[X_{3D} \;\; Y_{3D} \;\; Z_{3D}]^T$ in a 3D scene is $[x_{2D} \;\; y_{2D}]^T = [X_{3D} \;\; Y_{3D}]^T$.

[5] Epstein, Bontrager, and Park (1962) did not clearly specify shapes of triangles used in their experiments.

A triangle *ABC* is planar and a shape of *ABC* can be determined using a 2D coordinate system: $A = [0\ 0]^T$, $B = [1\ 0]^T$, and $C = [X_C\ Y_C]^T$. Any shape of *ABC* is characterized by the *x*- and *y*-coordinates x_C and y_C of *C*. The triangle *ABC* in the 3D scene can be written as:

$$T_\Delta + R_Z(\tau_\Delta)R_Y(\sigma_\Delta)R_Z(\rho_\Delta)\begin{pmatrix} A & B & C \\ 0 & 0 & 0 \end{pmatrix} \quad (1)$$

$$T_\Delta = \begin{bmatrix} X_\Delta \\ Y_\Delta \\ Z_\Delta \end{bmatrix}, R_Y(\theta_Y) = \begin{pmatrix} \cos\theta_Y & 0 & \sin\theta_Y \\ 0 & 1 & 0 \\ -\sin\theta_Y & 0 & \cos\theta_Y \end{pmatrix}, R_Z(\theta_Z) = \begin{pmatrix} \cos\theta_Z & -\sin\theta_Z & 0 \\ \sin\theta_Z & \cos\theta_Z & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where T_Δ represents a translation, R_Y and R_Z represent rotations about the *Y*- and *Z*- axes, and X_Δ, Y_Δ, Z_Δ, τ_Δ, σ_Δ, and ρ_Δ are free parameters. A 3D orientation of *ABC* is determined by slant (σ_Δ), tilt (τ_Δ), and roll (ρ_Δ) and its 3D position is determined by T_Δ. Then, a 2D orthographic projection with a uniform scaling of *ABC* in the 3D scene is:

$$s_\Delta \begin{bmatrix} X_\Delta \\ Y_\Delta \end{bmatrix} + s_\Delta R_{2D}(\tau_\Delta)\begin{pmatrix} \cos\sigma_\Delta & 0 \\ 0 & 1 \end{pmatrix} R_{2D}(\rho_\Delta)(A \quad B \quad C) \quad (2)$$

$$R_{2D}(\theta_{2D}) = \begin{pmatrix} \cos\theta_{2D} & -\sin\theta_{2D} \\ \sin\theta_{2D} & \cos\theta_{2D} \end{pmatrix}$$

where R_{2D} represents a 2D rotation on the image plane Π_I and s_Δ is a factor of the uniform scaling. Note that sign of s_Δ depends on where Π_I is for the perspective projection approximated by this orthographic projection. The sign of s_Δ is negative ($s_\Delta < 0$) if Π_I is behind of the center of projection relative to the 3D scene (e.g. retinae of eyes and sensors/films of real cameras). The sign is positive ($0 < s_\Delta$) if Π_I is in front of the center of projection (e.g. image screens of virtual cameras in many computer vision or computer graphics applications; see Figure 1 in the main text for an example). Equation (2) shows that the retinal image (the 2D orthographic projection with the uniform scaling) of *ABC* is an Affine transformation of *ABC*:

$$(a \quad b \quad c) = T_{2\times1} + M_{2\times2}(A \quad B \quad C) \quad (3)$$

$$T_{2\times1} = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}, M_{2\times2} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$$

where $a = [x_A\ y_A]^T$, $b = [x_B\ y_B]^T$, and $c = [x_C\ y_C]^T$ are the orthographic projections of *A*, *B*, and *C*.

Next, consider estimating the 3D position T_Δ and the 3D orientation (τ_Δ, σ_Δ, ρ_Δ) of the triangle *ABC* and the uniform scaling s_Δ from its retinal image *abc*. A shape of *ABC* and its projection *abc* are given. All the elements of $T_{2\times1}$ and $M_{2\times2}$ can be computed from *A*, *B*, *C*, *a*, *b*, and

c by solving Equation (3). The 3D position T_Δ can be estimated except for its Z-coordinate Z_Δ: $[X_\Delta\ Y_\Delta]^T = [t_1\ t_2]^T$. From Equations (2) and (3):

$$\begin{cases} m_{11} = s_\Delta \cos\sigma_\Delta \cos\rho_\Delta \cos\tau_\Delta - s_\Delta \sin\rho_\Delta \sin\tau_\Delta \\ m_{12} = -s_\Delta \cos\sigma_\Delta \sin\rho_\Delta \cos\tau_\Delta - s_\Delta \cos\rho_\Delta \sin\tau_\Delta \\ m_{21} = s_\Delta \cos\sigma_\Delta \cos\rho_\Delta \sin\tau_\Delta + s_\Delta \sin\rho_\Delta \cos\tau_\Delta \\ m_{22} = -s_\Delta \cos\sigma_\Delta \sin\rho_\Delta \sin\tau_\Delta + s_\Delta \cos\rho_\Delta \cos\tau_\Delta \end{cases} \quad (4)$$

Equation (4) can be re-written as:

$$\begin{cases} m_{22} + m_{11} = s_\Delta(1 + \cos\sigma_\Delta)\cos(\rho_\Delta + \tau_\Delta) \\ m_{21} - m_{12} = s_\Delta(1 + \cos\sigma_\Delta)\sin(\rho_\Delta + \tau_\Delta) \\ m_{22} - m_{11} = s_\Delta(1 - \cos\sigma_\Delta)\cos(\rho_\Delta - \tau_\Delta) \\ m_{21} + m_{12} = s_\Delta(1 - \cos\sigma_\Delta)\sin(\rho_\Delta - \tau_\Delta) \end{cases} \quad (5)$$

For a given sign of s_Δ, $\cos(\sigma_\Delta)$, τ_Δ, and ρ_Δ can be uniquely estimated by solving Equation (5).[6] A sign of the slant σ_Δ cannot be determined because $\cos(\sigma_\Delta) = \cos(-\sigma_\Delta)$. It is depth reversal ambiguity. From these facts, any Affine transformation $T_{2\times1}$ and $M_{2\times2}$ between *abc* and *ABC* can be decomposed into parameters that represent the 3D position and the 3D orientation of *ABC* with the depth reversal ambiguity. Namely, there are always 2 possible 3D interpretations of the given triangle *ABC* for the given retinal image *abc* under an orthographic projection.

References

Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, 39, 1143–1170.

Beck, J. & Gibson, J. J. (1955). The relation of apparent shape to apparent slant in the perception of objects. *Journal of Experimental Psychology*, 50, 125–133.

Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, 36, 1255–1264.

Campbell F. W. & Robson J.G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551–566.

Epstein, W., Bontrager, H., & Park, J. (1962). The induction of nonveridical slant and the perception of shape. *Journal of Experimental Psychology*, 63, 472–479.

[6] Note that $\tau_\Delta \pm 180°$ and $\rho_\Delta \pm 180°$ (any double sign) also satisfy Equation (5) if τ_Δ and ρ_Δ do. A 3D orientation specified by τ_Δ, σ_Δ, and ρ_Δ (see Equation 1) is identical with a 3D orientation specified by $\tau_\Delta \pm 180°$, σ_Δ, and $\rho_\Delta \pm 180°$ (any double sign).

Erkelens, C. J. (2015). Evidence for obliqueness of angles as a cue to planar surface slant found in extremely simple symmetrical shapes. *Symmetry*, 7, 241–254. https://doi.org/10.3390/sym7010241

Fischler, M. A. & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with apphcatlons to image analysis and automated cartography. *Communications of the ACM*, 24, 381–395.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, 87, 411–434.

Gallego, G., Mueggler, E., & Sturm, P. (2017, December 25). Translation of "Zur ermittlung eines objektes aus zwei perspektiven mit innerer orientierung" by Erwin Kruppa (1913). Retrieved from https://arxiv.org/abs/1801.01454

Gantz, L. & Bedell, H. E. (2011). Variation of stereothreshold with random-dot stereogram density. *Optometry and Vision Science*, 88, 1066–1071.

Gao, X., Hou, X., Tang, J., & Cheng, H. (2003). Complete solution classification for the perspective-three-point problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 930–943.

Gottheil, E. & Bitterman, M. E. (1951). The measurement of shape-constancy. *American Journal of Psychology*, 64, 406–408.

Griffiths, A. F. & Zaidi, Q. (2000). Perceptual assumption and projective distortions in a three-dimensional shape illusion. *Perception*, 29, 171–200.

Hartley, R. & Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge, MA: Cambridge University Press.

Howard, I. P. (2012). *Perceiving in Depth, Volume 3: Other Mechanisms of Depth Perception*. New York, NY: Oxford University Press.

Hu, X. & Ahuja, N. (1991). Motion estimation under orthographic projection. *IEEE Transactions on Robotics and Automation*, 7, 848–853.

Indow, T. (2004). *The Global Structure of Visual Space (Advanced Series on Mathematical Psychology, Vol. 1)*. Singapore: World Scientific Publishing.

Jayadevan, V., Sawada, T., Delp, E., & Pizlo, Z. (accepted). Perception of 3D symmetrical and nearly symmetrical shapes. *Symmetry*.

Kaneko, H. & Howard, I. P. (1997). Spatial limitation of vertical-size disparity processing. *Vision Research, 37*, 2871–2878.

Kruppa, E. (1913). Zur ermittlung eines objektes aus zwei perspektiven mit innerer orientierung [To determine a 3D object from two perspective views with known inner orientation]. *Sitzungsberichte der Mathematisch-Naturwissenschaftlichen Kaiserlichen Akademie der Wissenschaften*, 122, 1939–1948.

Li, S. & Xu, C. (2011). A stable direct solution of perspective-three-point problem. *International Journal of Pattern Recognition and Artificial Intelligence*, 25, 627–642.

Li, Y. & Pizlo, Z. (2011). Depth cues versus the simplicity principle in 3-D shape perception. *Topics in Cognitive Science*, 3, 667–685

Li, Y., Sawada, T., Shi, Y., Kwon, T., & Pizlo, Z. (2011). A Bayesian model of binocular perception of 3D mirror symmetric polyhedra. *Journal of Vision*, 11(4), 1–20.

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, 35, 549–568.

Mitsudo, H. (2007). Illusory depth induced by binocular torsional misalignment. *Vision Research*, 47, 1303–1314.

Mitsudo, H., Kaneko, H., & Nishida, S. (2009). Perceived depth of curved lines in the presence of cyclovergence. *Vision Research*, 49, 348–361.

Pei, S.-C. & Liou, L.-G. (1994). Finding the motion, position and orientation of a planar patch in 3D space from scaled-orthographic projection. *Pattern Recognition*, 27, 9–25.

Perkins, D. N. (1972). Visual discrimination between rectangular and nonrectangular parallelopipeds. *Perception & Psychophysics*, 12(3), 199–218.

Perkins, D. N. (1976). How good a bet is good form? *Perception*, 5(4), 396–400.

Pizlo, Z. (1994). A theory of shape constancy based on perspective invariants. *Vision Research*, 34, 1637–1658.

Pizlo, Z. (2008). *3D shape: Its unique place in visual perception*. Cambridge, MA: MIT Press.

Pizlo, Z. & Salach-Golyska, M. (1994). Is vision metric? Comment on Lappin and Love (1992). *Perception & Psychophysics*, 1994, 55, 230–234.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317, 314–319.

Sugihara, K. (1997). Three-dimensional realization of anomalous pictures – An application of picture interpretation theory to toy design. *Pattern Recognition*, 30(7), 1061–1067.

Sugihara, K. (2014a). A single solid that can generate two impossible motion illusions. *Perception*, 43, 1001–1005.

Sugihara, K. (2014b). Right-angle preference in impossible objects and impossible motion. In G. Greenfield, G. Hart, & R. Sarhangi (Eds.), *Proceedings of Bridges 2014: Mathematics, Music, Art, Architecture, Culture* (pp. 449–452). Phoenix, AZ: Tessellations Publishing.

Sugihara, K. (2016, August). *What Defeats Binocular Stereo?* Poster presented at the 39th European Conference on Visual Perception (ECVP), Barcelona, Spain.

Thompson, E. H. (1959). A rational algebraic formulation of the problem of relative orientation. *Photogrammetric Record*, 3, 152–159.

Wallach, H. & Moore, M. E. (1962). The role of slant in the perception of shape. *American Journal of Psychology*, 75, 289–293.

Watanabe, T. (1996). The estimation of the curvature of visual space with a visual triangle. *Japanese Journal of Psychology*, 67, 278–284.

Watt, R. J. (1984). Towards a general theory of the visual acuities for shape and spatial arrangement. *Vision Research*, 24, 1377–1386.

Zhang, Z.-L., Berends, E. M., & Schor, C. M. (2003). Thresholds for stereo-slant discrimination between spatially separated targets are influenced mainly by visual and memory factors but not oculomotor instability. *Journal of Vision*, 3, 710–724.

Vasily Minkov
Undergraduate Student, proveyourselfmail@gmail.com, School of Psychology, National Research University Higher School of Economics.

Tadamasa Sawada
Assistant Professor, PhD, tsawada@hse.ru (tada.masa.sawada@gmail.com), School of Psychology, National Research University Higher School of Economics.

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.