

# Testing a formal theory of perception is not easy: Comments on Yu, Todd & Petrov (2021) and Yu, Petrov & Todd (2021)

**Tadamasa Sawada**

School of Psychology, HSE University Moscow,  
Russian Federation



**Zygmunt Pizlo**

Department of Cognitive Sciences,  
University of California-Irvine Irvine, CA, USA



**Yu, Todd, and Petrov (2021, Journal of Vision) and their follow-up study (Yu, Petrov, & Todd, 2021, i-Perception) aimed at evaluating the role of three-dimensional (3D) symmetry in binocular shape perception by comparing their experimental data to predictions they derived from our computational models. We point out in this note that their predictions were incorrect, so their studies can neither reject nor support our models of 3D shape perception. We explain (1) the role of the data and the constraints in solving ill-posed inverse problems, (2) the role of binocular depth-order, as opposed to binocular depth-intervals in shape perception, (3) the nature and the effect of 3D compactness as an a priori constraint, and (4) the implications of the separation of binocular disparity and stereoacuity in the two functional streams in the visual cortex.**

In this note, we are commenting on the article by [Yu, Todd and Petrov \(2021\)](#) published in *Journal of Vision* and a follow-up article by [Yu, Petrov and Todd \(2021\)](#), published in *i-Perception*. We will refer to these articles as Yu1 and Yu2, respectively. These articles aimed at evaluating the role of three-dimensional (3D) symmetry in binocular shape perception. More specifically, they attempted to compare their experimental data to predictions that they derived from our computational models. Our models are based on several formal concepts taken from projective geometry, Inverse Problems Theory, symmetry, as well as the Regularization and Bayesian methods used to solve ill-posed problems. It seems that it is almost impossible to make any valid predictions of such models and to design meaningful experiments that test these models *without* proper attention to the underlying formalisms. We think that this was the main problem with the two studies that are discussed in this note.

1. Yu2 states in their abstract that “Prior theoretical analyses have shown that it is mathematically possible to compute the 3D shapes of symmetric stimuli [...], but those algorithms are useless for asymmetric objects.” The second part of this statement is true for objects that have no trace of symmetry whatsoever, such as [Rock and DiVita’s \(1987\)](#) wire stimuli or [Edelman and Bülthoff’s \(1992\)](#) bent paperclips. But note that Yu2 did not use such objects. Instead, they only considered objects that are approximately symmetrical, and they generated their asymmetrical objects by slightly distorting symmetrical objects. It is now widely recognized that an object’s symmetry is a continuous quantity that can vary between 0 (no symmetry at all) and 1 (perfect symmetry) ([Sawada, 2010](#); [Barlow & Reeves, 1979](#); [Tjan & Liu, 2005](#)). Objects that are approximately symmetrical can be handled naturally by the regularization and Bayesian methods used to solve ill-posed inverse problems. These methods combine the data with a priori constraints. The data are the retinal image(s), and the constraint in this case is 3D symmetry. If there is no solution that satisfies both exactly, the resulting percept may be closer to the data or closer to the constraints, depending on the reliability of each of these two components. So, contrary to Yu2’s claim, these methods can be applied to objects that are approximately symmetrical, but note that symmetrical objects allow for better reconstructions ([Sawada, 2010](#); [Jayadevan, Sawada, Delp, & Pizlo, 2018](#)).
2. Yu1 and Yu2 repeatedly refer to *binocular disparity*, which means to them one of the two main binocular capacities that allows an observer to estimate *depth intervals*, once the viewing distance is known or estimated. *Our theory of binocular shape perception makes no use of binocular depth intervals whatsoever* ([Li, Sawada, Shi, Kwon, & Pizlo, 2011](#); [Pizlo, Li, Sawada, & Steinman, 2014](#); [Jayadevan et al., 2018](#)). Instead, our theory uses *stereoacuity*, which is the ability to judge the *depth-order* of features ([Howard & Rogers, 1995](#)). Stereoacuity is a member of the visual *hyperacuties*. This hyperacuity does not produce systematic errors, and its random error is a fraction of the distance between neighboring receptors. In contrast, binocular depth intervals show large systematic as well as random errors. For example, at the viewing distance of 2 m, depth intervals are systematically underestimated by a factor of 2. Clearly, these two binocular capacities are very different. In fact, we know that they are analyzed in two different parts of the visual cortex (see point 4 below).



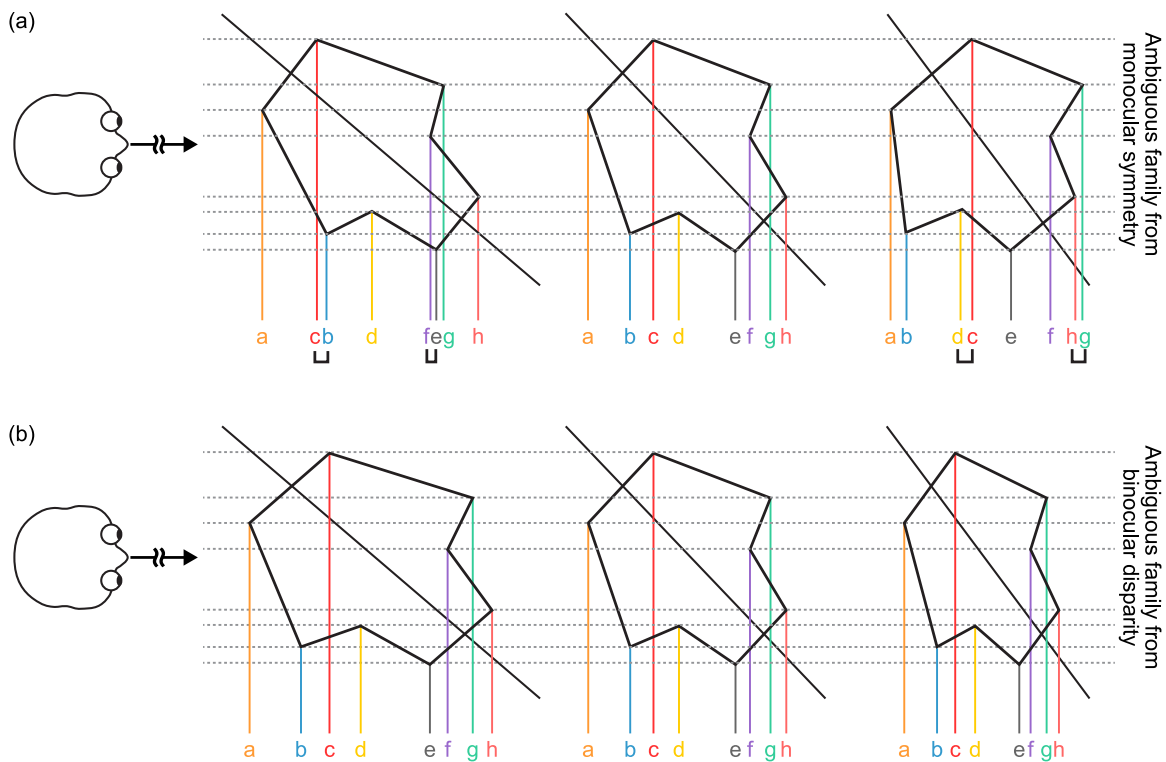


Figure 1. (a) An orthographic image of a 3D symmetrical shape determines this shape up to a one-parameter family represented by the slant of the symmetry plane (modified from Figure 5 in Pizlo et al., 2010). Veridical interpretation is possible by using information about the depth order of features, such as vertices of a polyhedron. (b) Depth order is constant and is useless when the only manipulation is a stretch or compression of the object along the depth direction (dotted lines). The horizontal dotted lines represent projecting lines in an orthographic projection, and the vertical colored lines represent the depths of vertices. Note that the depth order of the vertices changes frequently when the aspect ratio of the symmetrical shape changes (as shown in a).

Note that the way Yu1 and Yu2 designed and ran their experiments actually forced their subjects to use binocular *depth intervals* for their judgments. In each trial of their experiments, the observer was shown a pair of reference and adjustable stimuli that were stereoscopic images of 3D shapes. The reference shape was a volumetric polyhedron and the adjustable shape was the transformation of the reference shape. The transformation included uniform scaling, translation along the depth axis, and scaling along the depth axis. The observer adjusted depth scale to make the adjustable shape identical to the reference shape. Binocular *depth-order* was completely ineffective as a visual cue in this task because the adjustment of the depth scale changes *depth-intervals*, but not *depth-order* (see Figure 1b). So their experiments did not test *our* theory. Our binocular theory of shape perception capitalizes on the intersection of two constraints: *mirror-symmetry and depth order*, as illustrated in Figure 1a. It follows that our shape reconstruction model (both monocular and binocular) could not perform their task *unless we added a depth mechanism, based on depth intervals, to the existing shape mechanism* in our model. More precisely, our 2011 model would “perceive” a specific 3D reference shape in Yu1 experiment, namely a mirror symmetrical shape, but the task involving stretching or compressing the adjustable shape, could not possibly “measure” the model’s percept of the reference shape. Adding a depth mechanism based on depth intervals, to the adjustment, would help, except that adjusting the perceived depth interval by an observer results in biased and unreliable estimates (see point 4). *The only way to avoid using depth intervals when measuring 3D shape, is to view the shape from a different viewing direction*, the procedure universally used in shape constancy experiments for more than half a century. The reported poor performance of their subjects, including large individual variability, can be attributed to this problem: the subjects, according to our model, saw the correct 3D symmetrical shape, but the adjustment tool they were offered was useless for the purpose of adjusting the perceived shape, so, the subjects had to rely on binocular depth intervals. This is actually illustrated well by their examples. Our model cannot tell the difference between the three shapes in Figure 11 in Yu1 or between the two shapes in Figure 5 in Yu2, unless we add binocular depth perception to our shape model. Note that a reader will have a hard time making out this difference, as well.

But we want to be clear what we really mean when we say “adding binocular depth perception to our shape model.” Our model recovers 3D shapes by minimizing a cost function that includes symmetry, compactness and

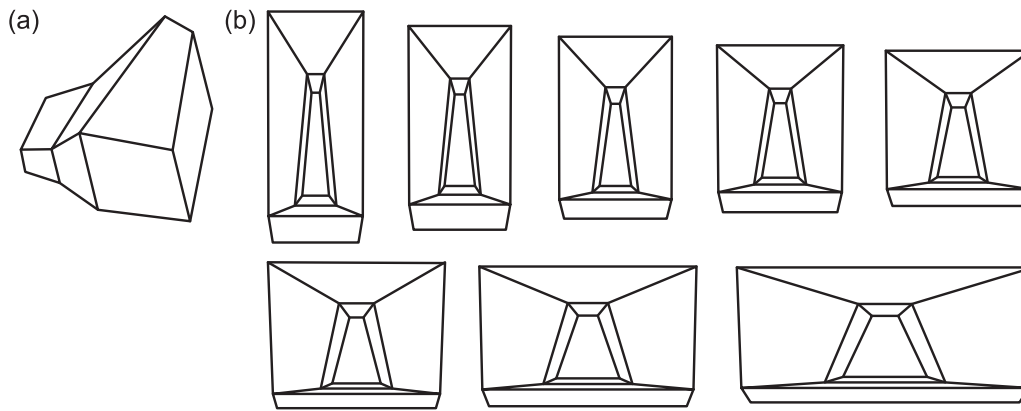


Figure 2. Frontal views (b) of eight shapes illustrated in demo 2.3 that accompanies Pizlo et al.'s (2014) book (<http://shapebook.psych.purdue.edu/2.3/>). Any of these eight shapes could have been a reference shape that produced the orthographic image shown in (a), but only one of them (shown on top-right) is maximally compact. Maximizing 3D compactness can produce a veridical interpretation only if the reference shape is maximally compact.

stereoacuity. This model would perceive both the reference and adjustable 3D shapes in Yu1's experiment as symmetrical, but when asked to match the depth ranges of these two shapes, the model would have to *switch* to binocular depth intervals. As we indicated in our 2011 paper (Li et al., 2011, p.13) adding binocular depth intervals to the cost function would lead to biased and unreliable shape perception. We did not see large biases of the 3D shape perception in the performance of our subjects (except for one subject who did not have stereoacuity). Furthermore, our binocular model could account very well for our subjects' performance without using depth intervals. So, we don't think that shape perception and depth perception should be combined. It is better to keep them separate. It follows that when the subjects in Yu1's and Yu2's experiments could not rely on adjusting 3D shapes, they *switched their attention* to depth.

To summarize this point, we think that there are two different mechanisms in the visual system, one underlying shape and the other underlying depth perception. Yu1 and Yu2 propose that there is one mechanism based on depth perception. In addition, our model also includes a monocular observer as a special case. Their model does not have a monocular observer.

3. Maximizing 3D compactness, without any other components in the cost function that is used by the visual system, *could not* produce veridical performance in our experiments, contrary to Yu1's and Yu2's claims, because we varied the compactness of our reference objects by changing the objects' aspect ratios over a range of 1/5 to 5, a factor of 25 (Li et al., 2011). We used this range of aspect ratios with 1000 randomly generated polyhedra, and computed that compactness changed by a factor of 3.4 (with standard deviation 0.8) (Figure 2 shows 8 possible reference shapes with different aspect ratios and different compactness). We think that Yu1 and Yu2 overlooked the fact that we manipulated 3D compactness of the reference objects. Our explanation of this point can be easily verified, simply by examining the performance of our monocular model in Li et al. (2011) (see also Li et al., 2009). This model took a 2D image and chose the *maximally compact* 3D shape from its family of 3D symmetrical shapes (to be precise, our model used a modified 3D compactness, but note that this is not critical for the present argument). Performance of the monocular model was not perfect, precisely because the model always produced maximally compact objects, while the reference objects were not necessarily maximally compact. The only way Yu1's subjects could have produced a veridical match of an adjustable shape to an invisible reference shape (see their Fig. 12), is if the authors used compact reference shapes. *Maximal compactness* can produce veridical percepts trivially with *maximally compact* reference objects.
4. Finally, look at Figure 1 in Yu2. Every observer can easily figure out which shape is symmetrical and which is not, and the perceptual result is the same in both monocular and binocular viewing. So, shape perception is actually very good—we can detect even 10% shifts of the vertices in 3D. You don't need depth intervals to explain this example. Depth intervals produced by binocular disparity are actually confounding the entire problem. The binocular 3D shape percept, according to our theory, is produced without using depth intervals simply because they *are not needed* to produce a veridical percept, and if depth intervals had been included, they would have distorted the perception of the shape (see point 2 above). Binocular 3D shape perception in our theory is produced by using depth-order information, which is known as stereoacuity. It is worth pointing out that stereoacuity is processed in the ventral stream of the visual brain, which also processes shape information, whereas the dorsal stream processes coarse depth information based on binocular disparity (Yoshioka, Doi, Abdolrahmani, & Fujita, 2021; Verhoef, Vogels, & Janssen, 2016). It is therefore reasonable to assume that an observer can recover

the veridical 3D shape of an object on the basis of stereoacuity processed in the ventral stream although the observer can still produce a biased perception of depth intervals that are processed in the dorsal stream. In plain English, the experiments described by Yu1 and Yu2 are examples of using a biased yardstick to measure a perfect object. The result of the measurement is biased but not because there is a problem with the object that is being measured, but because the measuring tool is biased.

In conclusion, the studies described by Yu1 and Yu2 neither reject nor support our theory of 3D shape perception. Their studies do add to the existing literature on systematic distortions of binocular depth perception, but the influence of their studies is likely to be weaker than it might have been because of the problems we described above.

*Keywords:* 3D shape perception, symmetry, stereoacuity, inverse problems, compactness

## Acknowledgments

Commercial relationships: none.

Corresponding author: Tadamasa Sawada.

Email: tada.masa.sawada@gmail.com, tsawada@hse.ru.

Address: HSE University, Krivokolenny per., 3, Moscow, Russia.

## References

- Barlow, H. B., & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, *19*, 783–793.
- Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3D objects. *Vision Research*, *32*, 2385–2400.
- Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford: Oxford University Press.
- Jayadevan, V., Sawada, T., Delp, E., & Pizlo, Z. (2018). Perception of 3D symmetrical and nearly symmetrical shapes. *Symmetry*, *10*(344), 1–24.
- Li, Y., Pizlo, Z., & Steinman, R. M. (2009). A computational model that recovers the 3D shape of an object from a single 2D retinal representation. *Vision Research*, *49*(9), 979–991.
- 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.
- Pizlo, Z., Li, Y., Sawada, T., & Steinman, R. M. (2014). *Making a machine that sees like us*. Oxford: Oxford University Press.
- Pizlo, Z., Sawada, T., Li, Y., Kropatsch, W. G., & Steinman, R. M. (2010). New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume. *Vision Research*, *50*(1), 1–11.
- Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, *19*, 280–293.
- Sawada, T. (2010). Visual detection of symmetry in 3D shapes. *Journal of Vision*, *10*(6), 1–22.
- Tjan, B. S., & Liu, Z. (2005). Symmetry impedes symmetry discrimination. *Journal of Vision*, *5*, 888–900.
- Verhoef, B.-E., Vogels, R., & Janssen, P. (2016). Binocular depth processing in the ventral visual pathway. *Philosophical Transactions of the Royal Society B*, *371*, 20150259.
- Yoshioka, T.W., Doi, T., Abdolrahmani, M., & Fujita, I. (2021). Specialized contributions of mid-tier stages of dorsal and ventral pathways to stereoscopic processing in macaque. *eLife*, *10*, e58749.
- Yu, Y., Petrov, A. A., & Todd, J. T. (2021). Bilateral symmetry has no effect on stereoscopic shape judgments. *i-Perception*, *12*(4), 2041669521104-2644.
- Yu, Y., Todd, J. T., & Petrov, A. A. (2021). Failures of stereoscopic shape constancy over changes of viewing distance and size for bilaterally symmetric polyhedra. *Journal of Vision*, *21*(6):5, 1–19.