

Are mirror-symmetric objects of special importance for 3D shape perception? A reply to Sawada and Pizlo (2022)

Alexander A. Petrov

Department of Psychology, The Ohio State University,
Columbus, OH, USA



Ying Yu

Department of Psychology, The Ohio State University,
Columbus, OH, USA



James T. Todd

Department of Psychology, The Ohio State University,
Columbus, OH, USA



Yu, Todd, and Petrov (2021) and Yu, Petrov, and Todd (2021) investigated failures of shape constancy that occur when objects are viewed stereoscopically at different distances. Although this result has been reported previously with simple objects such as pyramids or cylinders, we examined more complex objects with bilateral symmetry to test the claim by Li, Sawada, Shi, Kwon, and Pizlo (2011) that the perception of those objects is veridical. Sawada and Pizlo (2022) offer several criticisms of our experiments, but they seem to suggest that the concept of shape is defined by what is computable by their model. If stimuli are used that cannot be discriminated by their model, they are dismissed as degenerate, and tasks that cannot be performed by their model are assumed to be based on something other than shape. This allows them to disregard empirical evidence that is inconsistent with their model. We argue, in contrast, that all reliable aspects of shape perception are deserving of explanation. We also argue that there are many different attributes of shape and many different sources of information about shape that may be relevant in different contexts. It is unlikely that all of them can be explained by a single model.

Introduction

One of the central goals of perceptual theory is to develop computational models for computing three-dimensional (3D) shape from visual information. This work began in the 1970s with the pioneering research of Horn (1975) on the analysis of 3D shape from shading, and the related work of Ullman (1979) on the analysis of 3D shape from motion. Because the mapping between the physical environment and optical projections is many to one, the general approach used by all such models is to assume some regularizing

constraints on the environment to limit the number of possible interpretations. For example, in the analysis of 3D shape from shading it is typically assumed that an object reflects light uniformly in all directions and that it is illuminated from a single direction. Similarly, in the analysis of 3D shape from motion it is typically assumed that the object is moving rigidly relative to the observer.

These assumed constraints can often reveal important limitations of a computational model. Although the use of constraints is mathematically unavoidable, the resulting analyses may be of limited use if their underlying assumptions are frequently violated in the natural environment. With respect to human perception, a good model should be able to produce accurate estimates of 3D shape in all conditions where human judgments of 3D shape are accurate, but it should also produce systematic distortions of estimated shape in all conditions where the perception of 3D shape is systematically distorted.

Like all computational models of 3D shape perception, the one developed by Pizlo and colleagues (Li, Pizlo, & Steinman, 2009; Li, Sawada, Shi, Kwon, & Pizlo, 2011) makes a number of assumptions that limit the scope of its applicability. It optimizes an objective function that includes terms related to the symmetry, planarity, and compactness of the estimated object. The object must have at least four visible pairs of corresponding points on opposite sides of the plane of 3D bilateral symmetry, and the object must be oriented so that its symmetry plane is neither parallel nor perpendicular to the observer's line of sight.

This type of special-purpose mechanism could only be useful in those relatively rare instances when its assumed constraints are satisfied. However, Pizlo and colleagues have marketed their approach as a general theory of shape perception. The rhetorical device they

Citation: Petrov, A. A., Yu, Y., & Todd, J. T. (2022). Are mirror-symmetric objects of special importance for 3D shape perception? A reply to Sawada and Pizlo (2022). *Journal of Vision*, 22(4):16, 1–5, <https://doi.org/10.1167/jov.22.4.16>.



use to achieve this is to implicitly define the concept of shape as that which is computable by their model. Anything that does not satisfy the assumptions of their model is labeled as “degenerate” and can therefore be ignored, as shown in the following quotation from [Pizlo, Sawada, Li, Kropatsch, and Steinman \(2010\)](#):

“... few actually complex 3D shapes have been used to study shape perception during the 30 years since Gibson’s death, and even these shapes have tended to be too simple, e.g., elliptical cylinders and rectangular pyramids. Furthermore, these 3D shapes were not only too simple to be used in studies of shape, they were almost always presented from very special viewing directions, called ‘degenerate views’, that is, views specifically chosen to remove all 3D shape information ... It is not surprising, then, that the shape judgments in these experiments were very variable, as well as biased” (p. 3).

The purpose of this argument is to dismiss any evidence that may challenge these authors’ claim that the perception of 3D metric structure is veridical, but it is useful to consider some of the actual stimuli they are labeling as “degenerate.” [Figure 1](#) shows stereograms of two simple objects at “degenerate” orientations. As is clear from the previous quotation, Pizlo and colleagues contend that the use of such stimuli intentionally removes “all 3D shape information.” However, readers who can free-fuse will quickly recognize the top panel of this figure as a 3D pyramid and the bottom panel as a 3D cylinder. [Pizlo et al. \(2010\)](#) are correct that perceptions of these stimuli are often biased, but they misrepresent the literature with respect to the

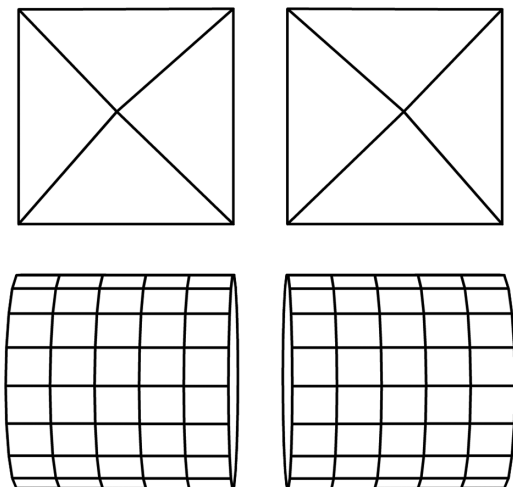


Figure 1. Stereograms of a pyramid (top) and a cylinder (bottom) at a fronto-parallel orientation. The model by Pizlo and colleagues is unable to estimate the shapes of these objects because they do not satisfy its underlying assumptions. They therefore label them as “degenerate.” However, human observers with functioning stereo vision can identify these objects quite easily.

reliability of observers’ judgments. When objects like this are presented at different distances in depth, the apparent depth-to-width ratios become systematically compressed as viewing distance is increased, and this result has been replicated in dozens of experiments over the past 100 years (for a review, see [Todd & Norman, 2003](#)). Pizlo and colleagues consider the use of “degenerate” stimuli to be a flaw in the design of these studies because their model is incapable of analyzing these stimuli. We consider it to be a flaw in their model, because it cannot explain a well-documented finding in the literature on stereoscopic shape perception involving simple 3D objects that are easily recognizable by all observers.

Our recent experiments ([Yu, Petrov, & Todd, 2021](#); [Yu, Todd, & Petrov, 2021](#)) were designed to examine if the use of more complex stimuli at different orientations would have any effect on the systematic distortions of apparent shape caused by changes in viewing distance that have been reported in previous investigations. According to [Pizlo and de Barros \(2021\)](#): “When an object is mirror-symmetrical, shape constancy is perfect, or nearly so” (p. 14). The stimuli we used were quite similar to those employed in Pizlo’s studies: They were mirror-symmetrical; they were sufficiently complex; and they were presented at “non-degenerate” orientations. Nevertheless, our empirical results revealed exactly the same patterns of perceptual distortion over changes in viewing distance that have been reported in previous studies with “degenerate” stimuli.

In their criticism of this research, [Sawada and Pizlo \(2022\)](#) claim that:

“... it is almost impossible to make any valid predictions of such [computational] models [of 3D shape perception] 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” (p. 1).

It is important to emphasize that shape constancy is a mechanism-independent concept, and so are the necessary conditions for establishing constancy and the sufficient conditions for demonstrating failures of constancy. It is quite possible to treat computational models (and the human visual system, for that matter) as black boxes and test them solely on the basis of their inputs and outputs. In particular, if physically different shapes are perceived to be the same, then shape constancy fails. Similarly, if physically identical shapes are perceived to be different, then shape constancy also fails. The mechanisms inside the black box are irrelevant.

Another criticism raised by [Pizlo and de Barros \(2021\)](#) concerns the matching task we employed:

“Why are so many contemporary students of vision still arguing whether shape constancy is real? The main reason (although not the only reason) for the existing confusion surrounding shape constancy is the fact that shape constancy has not been treated as a perceptual invariant related to a group of transformations. But the only way to make sure that you are studying shape as an invariant of rigid translation and rotation is to have the subject look at the object after it has been subjected to rigid translation and rotation. ... The bottom line is as follows: the 3D shape is invariant under 3D rigid motion and must be tested as such. The subject must be shown a 3D object from more than one 3D viewing direction in order to verify whether the perceived shape itself is invariant under 3D rigid motion” (p. 6).

This argument is invalid. Yes, it is indeed necessary to demonstrate shape invariance with respect to *all* members of the relevant group of transformations in order to *establish* shape constancy. For metric shape, this is the similarity group, which includes all translations, rotations, and uniform scaling transformations. However, to demonstrate *failures* of shape constancy it is sufficient to demonstrate that perceived 3D shape varies systematically under *any* of these transformations. Our experiments revealed that apparent 3D shape varies systematically with respect to translation in depth and uniform scaling. Each of these results constitutes a clear violation of shape constancy.

[Sawada and Pizlo \(2022\)](#) also raise a concern that:

“Our model cannot tell the difference between the three shapes in Figure 11 in ([Yu et al., 2021](#)) or between the two shapes in Figure 5 in ([Yu et al., 2021](#)), unless we add binocular depth perception to our shape model. Note that the reader will have a hard time making out this difference as well” (p. 2).

Given the uncontested fact that these figures depict objects with different metric shapes, the conclusion that we draw from these observations is that neither Pizlo’s model nor the human visual system achieves shape constancy in these instances. When observers perform the adjustment task in these experiments, the changes in the depth-to-width aspect ratios of the stimuli are clearly visible. We pressed the observers on this point in our initial instructions and the debriefing to ensure this was the case. If Pizlo’s model cannot detect these changes, which are clearly visible to the observers, then that is a weakness of the model, not a flaw in our experimental design. Note how their criticism contains an implicit suggestion that shape is defined by the capabilities of their model, and, consequently, that our experimental task cannot possibly be about shape because it relies heavily on distinctions that are invisible to their model.

[Sawada and Pizlo \(2022\)](#) raise a second concern about our experimental task:

“The binocular 3D shape percept, according to [[Sawada and Pizlo’s](#)] 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” (p. 3).

They also argue that “3D shape perception” and “depth perception” are two distinct tasks that must not be confused. According to this argument, their model analyzes the former, whereas our experimental procedure tested the latter.

There are several aspects of these comments that deserve to be rebutted. First, the task employed in our studies could not have been performed by simply comparing depth intervals, because the objects to be compared were always of different sizes. Shape judgments in that context require a comparison of the aspect ratios. Second, the model developed by [Li et al. \(2009, 2011\)](#) explicitly computes the Cartesian coordinates (x, y, z) of every visible vertex on an object. The depth interval of an object can thus be determined trivially by the difference between the largest and smallest value of z . Why, then, should judgments of distance intervals be considered as something independent from the analysis of shape? This sounds like another convenient excuse to dismiss an empirical result that is incompatible with their model. Finally, Pizlo and colleagues have stated on numerous occasions (including the previous quotation) that shape perception is veridical for objects that satisfy the conditions of their model. If the use of binocular disparities somehow contaminates the computation of 3D shape, then why wouldn’t observers simply ignore that information and base their judgments solely on symmetry and depth order? The answer to this question is quite simple: The variations in 3D shape that observers were asked to evaluate are undetectable by Pizlo’s model, although they were quite noticeable to the observers. Note that there is a persistent theme in these arguments. If shape is defined by what can be computed by their model, then any violations of constancy or veridicality cannot possibly involve shape perception. This is a very convenient rhetorical device for dismissing contradictory evidence.

At the end of the day, our criticisms of the methods of Pizlo and his colleagues and their criticisms of our methods are likely to be irrelevant. What will matter most in evaluating their model is the breadth of phenomena (or lack thereof) that it is able to explain. Their model is designed specifically for bilaterally symmetric (or nearly symmetric) polyhedra of sufficient structural complexity. Although many objects we encounter in the environment satisfy these criteria, there are many more that do not. If their model cannot handle simple shapes like the ones shown in [Figure 1](#) that are easily identified by anyone, then that is a serious

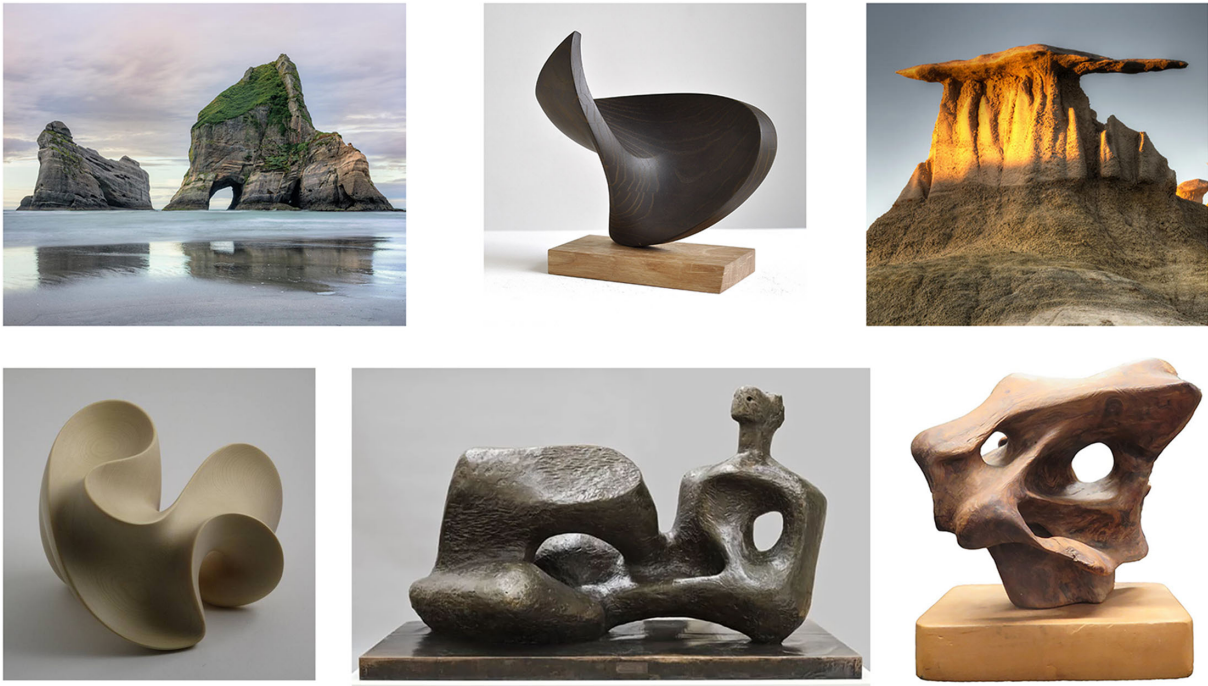


Figure 2. Images of abstract sculptures and natural rock formations that do not satisfy the underlying assumptions of Pizlo and colleague's model. Almost all observers report that these images provide compelling perceptions of 3D shape.

problem for what they are proposing as a model of human perception.

To better appreciate the wide range of objects excluded from their analysis, it is useful to consider the images in [Figure 2](#). They depict abstract sculptures and natural rock formations, none of which satisfies the underlying assumptions of their model. Should we conclude that these objects do not possess the property of “shape” as argued by Pizlo and colleagues? We suspect that most readers will quickly recognize that argument as an obvious attempt to paper over a serious shortcoming of their model. Almost all observers agree that each of these images produces a compelling perceptual appearance of 3D shape and that they are stunningly beautiful. A complete theory of shape perception should be able to account for the perceptual appearance of these objects as well as plane-faced polyhedra.

Our overall position is that it is unlikely that any single model can account for all aspects of shape perception. Indeed, there is considerable evidence that the human visual system has many special-purpose modules for determining 3D shape from different types of visual information, such as shading, texture, motion, or binocular disparity. Perhaps Pizlo's model could be considered as one component within that framework, although its inability to cope with simple basic shapes such as pyramids or cylinders is problematic.

In a recent review article on the concept of shape within multiple fields ([Todd & Petrov, 2022](#)) we

discuss numerous models for how 3D shapes might be perceptually represented. The model proposed by Pizlo and colleagues is an outlier because it is focused primarily on Euclidean metric structure. It is also unusual because these authors insist that the perception of 3D metric structure is veridical, despite the fact that this claim is inconsistent with the vast majority of psychophysical experiments that have explored this issue.

[Todd and Petrov \(2022\)](#) argue that shape is not a unitary property but rather a collection of many object attributes, some of which are more perceptually salient than others. Because the relative importance of these attributes can be context dependent, there is no obvious single definition of shape that is universally applicable in all situations. Whereas the metric properties of Euclidean geometry may be of paramount importance to a tool and die maker, they are largely irrelevant to a biologist who is trying to classify the biological forms of different species. There is considerable evidence to suggest that the most perceptually salient aspects of shape are those that involve affine, projective, and topological properties and that Euclidean metric structure is of relatively minor importance. The problem with the theory proposed by Pizlo and colleagues is that it focuses on the minor aspects of shape while ignoring the more significant ones, and their arguments twist into knots when trying to evade the large body of empirical evidence that is inconsistent with their position.

Keywords: shape constancy, computational model, methodology, structure from stereo, structure from symmetry

Acknowledgments

Supported by a grant from the National Science Foundation (BCS-1849418).

Commercial relationships: none.

Corresponding author: Alexander A. Petrov.

Email: apetrov@alexpetrov.com.

Address: Department of Psychology, The Ohio State University, Columbus, OH, USA.

References

- Horn, B. (1975). Obtaining shape from shading information. In P. Winston (Ed.), *The psychology of computer vision* (pp. 115–155). New York: McGraw-Hill.
- 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, <https://doi.org/10.1016/j.visres.2008.05.013>.
- 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):11, 1–20, <https://doi.org/10.1167/11.4.11>.
- Pizlo, Z., & de Barros, J. A. (2021). The concept of symmetry and the theory of perception. *Frontiers in Computational Neuroscience*, *15*, 681162, <https://doi.org/10.3389/fncom.2021.681162>.
- 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, <https://doi.org/10.1016/j.visres.2009.09.024>.
- Sawada, T., & Pizlo, Z. (2022). Testing a formal theory of perception is not easy: Comments on Yu, Todd, & Petrov (2021) and Yu, Petrov, & Todd (2021). *Journal of Vision*, *22*(4):15, 1–4, <https://doi.org/10.1167/jov.22.4.15>.
- Todd, J. T., & Norman, J. F. N. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception & Psychophysics*, *65*(1), 31–47, <https://doi.org/10.3758/BF03194781>.
- Todd, J. T., & Petrov, A. A. (2022). The many facets of shape. *Journal of Vision*, *22*(1):1, 1–30, <https://doi.org/10.1167/jov.22.1.1>.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Yu, Y., Petrov, A. A., & Todd, J. T. (2021). Bilateral symmetry has no effect on stereoscopic shape judgments. *i-Perception*, *12*(4), 1–17, <https://doi.org/10.1177/20416695211042644>.
- 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, <https://doi.org/10.1167/jov.21.6.5>.