Title
Curve-Based Shape Representation in Visual Perception

Author
Baker, Nicholas

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Curve-Based Shape Representation in Visual Perception

A dissertation submitted in partial satisfaction

of the requirement for the degree

Doctor of Philosophy in Psychology

by

Nicholas Sayres Baker

2020

ABSTRACT OF THE DISSERTATION

Curve-Based Shape Representation in Visual Perception By

Nicholas Sayres Baker

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2020

Professor Philip Kellman, Chair

Shape is the predominant cue for object recognition in visual perception. Though many studies have demonstrated the psychological importance of shape information, much remains unknown about how the visual system forms representations of shape. Shape representations are unlikely to be a literal recording of an object's boundary. Rather, representations of shape are abstract in that they encode relations between parts, are economical, selectively encoding information present in the physical stimulus, and are invariant to 2D transformations and changes to the properties of local elements.

In this dissertation, I examine evidence for the theory that representations of shapes are formed by partitioning a contour into regions of similar curvature and representing segments with a single curvature value. I first develop a computational model for how contours could be recoded abstractly as sets of constant curvature segments. I first experimentally tested two free parameters in the model and then tested the model's ability to predict the perceptual difference between pairs of shapes. In Chapter 2, I showed how the visual system could encode constant curvature representations of shape from activations of oriented luminance contrast detectors in early vision, bridging a theoretical gap between subsymbolic activations that are responsive to light energy and symbolic representations that are concerned with objects, contours, and surfaces.

In Chapters 3 and 4, I applied the constant curvature theory to two interesting domains of shape perception. First, I tested how and why people encode shape representations from arrays of unconnected dots. Consistent with the constant curvature theory of shape, dot arrays that were perceived to have curvilinear contours were more easily represented as shapes than dot arrays perceived to have straight edges joined at corners. In Chapter 4, I studied shapes with both global form and high frequency local contour features. Evidence was found for a hypothesis that local and global contour features are encoded independently and in separate systems. In this theory, global features are extracted from large curvature detectors and described in detail while local contour features are extracted from small curvature detectors and encoded with a few descriptive statistics rather than as individual features.

The dissertation of Nicholas Sayres Baker is approved.

Tao Gao

Hongjing Lu

Zili Liu

Philip Kellman, Committee Chair

University of California, Los Angeles

2020

For my parents, whose love, wisdom, and encouragement I will always be grateful for and for my Heavenly Father, a constant source of peace, joy and hope

Table of Contents

Acknowledgements

First and foremost, I would like to thank my advisor, Phil Kellman, for his kind and patient mentorship throughout my PhD work. Special thanks also to Pat Garrigan, upon whose research much of this work is built. I heartfeltly thank my committee and the Human Perception Lab, whose feedback and insights have been essential to this work. Finally, thank you to my friends and family for your encouragement and support.

Vita

Nicholas Baker
Start of PhD: September 2014

Education
B.A. in Cognitive Science from Johns Hopkins University, with honors
M.A., Computational Cognition from University of California, Los Angeles

Fellowships and Awards
UCLA Graduate Dean's Scholarship
Graduate Student Research Mentorship
National Science Foundation Research Traineeship for ModEling and uNdersTanding human behaviOR (MENTOR)
UCLA Dissertation Year Fellowship

Papers under Review
Baker, N., Garrigan, P., & Kellman, P.J. (2020). Evidence for constant curvature primitives in abstract shape representations. *JEP: General*.

Papers in Preparation
Baker, N. & Kellman, P.J. (2020). Constant curvature modeling of abstract shape representation.

Baker, N. & Kellman, P.J. (2020). Why do we perceive shapes in unconnected dots?

Publications
Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2020). Local features and global shape information in object classification by deep convolutional neural networks. *Vision Research*, Accepted 6 April, 2020.

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS computational biology*, *14*(12), e1006613.

Baker, N., Kellman, P. J., Erlikhman, G., & Lu, H. (2018). Deep convolutional networks do not perceive illusory contours. In T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1310-1315). Austin, TX: Cognitive Science Society.

Baker, N., & Kellman, P. J. (2018). Abstract shape representation in human visual perception. *Journal of Experimental Psychology: General*, *147*(9), 1295.

Thomas V Papathomas, Nicholas Baker, Arielle S Yeshua, Xiaohua Zhuang, Andrew Ng. The Ingenious Mr Hughes: Combining forced, flat, and reverse perspective all in one art piece to pit objects against surface. *i-Perception*. 01/2012; 3(3):182-7.

Colloquia
"Shape perception in human vision and deep convolutional networks." Cognitive Forum at the University of California, Los Angeles, March 8, 2019.

"Shape perception in human vision and deep convolutional networks." Invited talk at École Polytechnique Fédérale de Lausanne, March 26, 2020.

Conference Presentations

Baker, N.*, & Kellman, P.J. (2019). *Constant curvature modeling of abstract shape representation*. Talk at the Nineteenth Annual Meeting of Vision Sciences Society, St. Pete's Beach, FL.

Kellman, P. J., Erlikhman, G., Baker, N.*, & Lu, H. (2019). *Recursive Networks Reveal Illusory Contour Classification Images.* Poster at the Nineteenth Annual Meeting of Vision Sciences Society, St. Pete's Beach, FL.

Baker, N.*, Erlikhman, G., Kellman, P., & Lu, H. (2018). *Deep convolutional networks do not perceive illusory contours*. Poster at the 40th Annual Meeting of the Cognitive Science Society, Madison, WI.

Baker, N.*, Lu, H., Erlikhman, G., & Kellman, P. J. (2018). *Deep convolutional networks do not classify based on global object shape.* Talk at the Eighteenth Annual Meeting of Vision Sciences Society, St. Pete's Beach, FL.

Kellman, P.J.* & Baker, N (2018). *Visual perception of shape in humans and deep convolutional neural networks*. Talk at the Society of Experimental Psychologists 2018, Tucson, Arizona.

Baker, N.*, Lu, H., Erlikhman, G., & Kellman, P. J. (2017). *Deep convolutional networks do not classify based on global object shape.* Poster at the 2017 Annual Meeting of the Psychonomics Society, Vancouver, Canada.

Baker, N.*, Lu, H., Erlikhman, G., & Kellman, P. J. (2017). *Deep convolutional networks do not classify based on global object shape.* Talk at the 2017 Annual Meeting of the Configural Processing Consortium, Vancouver, Canada.

Baker, N.*, Erlikhman, G., & Kellman, P. J., & Lu, H. (2017). *Classification images reveal that deep learning networks fail to perceive illusory contours.* Talk at the Seventeenth Annual Meeting of Vision Sciences Society, St. Pete's Beach, FL.

Baker, N.* & Kellman, P.J. (2017). *Psychophysical investigations into skeletal shape representations.* Poster at the Seventeenth Annual Meeting of Vision Sciences Society, St. Pete's Beach, FL.

Baker, N.* & Kellman, P. J. (2017). *Abstract shape representation.* Talk at the 2016 Annual Meeting of the Configural Processing Consortium, Boston, MA.

Baker, N.* & Kellman, P. J. (2016) T*emporal properties of abstract shape perception*. Poster at the Sixteenth Annual Meeting of Vision Sciences Society, St. Pete's Beach, FL.

Thomas V. Papathomas, Marilena Karakatsani*, Steven M. Silverstein, Nicholas Baker, Marcel de Heer. "Experiments and Computational Models for the Ames Window Illusion". European Conference of Visual Perception 2013, 25 - 29 Aug 2013, P7-5.

*Indicates the presenting author

Introduction

In this research, I put forward theoretical and empirical findings that aim to clarify how the shape of objects is perceived and represented in human vision. Many of the questions addressed in this document harken back to century-old insights from Gestalt psychologists, who emphasized that shape is a relational notion, and that a shape percept is largely divorced from the local elements that induce it (Wertheimer, 1923; Koffka, 1935). I begin from the premise that shape is relational, and that what we encode about an object's shape is qualitatively different than the physical properties of its bounding contour. In this document, I am primarily concerned with two fundamental questions of shape perception. First, how does the visual system form a compact description of shape that is invariant to changes in size, orientation, position, or local elements? Second, how does the predominance of shape in human vision influence our perception of stimuli with no inherent structure like arrays of unconnected dots or allow us to find equivalence between objects with very different local elements but similar global form such as when we see objects in clouds.

Implicit in the phenomenology and perceptual processing of shapes, is the idea that shape representations are *abstract*. Abstraction is a complicated notion, having several related meanings (see Barsalou & Weimer-Hastings, 2005, for useful discussion). In our use of "abstract" in the realm of shape perception, we intend three related ideas. The first is encompassed by the idea above that shape is a relational notion, i.e, by abstract, we mean at least relevant information consists in relationships defined over, but not by, lower order constituents; such relationships can be described as binding the value of a variable (Marcus, 2001; Overlan, Jacobs & Piantadosi, 2017). For instance, to be a square does not mean that a side of the form has to be of a certain length, but that the length of one side must equal the length of any other

side, formally expressible as, for any two sides a and b, length(a) is equal to length(b). Intuitively, we recognize shape representations as abstract when we notice that a cloud appears to resemble a dog, or when we notice that two objects of different sizes, composition, and orientation share the same shape. Our perceptual representations of shape allow these "matching" experiences despite radically different contexts or constituent elements. A second, related, idea is that at least some shape descriptions, including, arguably, those used in the brain, capture information economically (Attneave, 1954); they comprise a summary description from which much specific stimulus information has been discarded. A third aspect of abstraction in human shape representation is more or less the converse: the representation is abstract in adding something that was not present in the stimulus. Baker & Kellman (2018) studied abstract shape perception experimentally using separated dot elements around virtual contours of unfamiliar, smooth, closed 2D shapes. We found that, beyond a very brief interval after stimulus offset, encoding of specific elements was poor or non-existent. In contrast, shape representations were encoded that supported accurate same / different judgments, across displays, despite transformations of position, orientation, and scale. Such results imply a representation that has captured relations among the inputs while discarding the concrete values of the inputs. Moreover, in these studies, no continuous contour shape information was actually given in the stimulus. The dots used in each display could have been connected in a virtually unlimited number of ways (or not at all). The particular shape representations that supported task performance were abstract in a) being derived from relations; b) being more economical descriptions in that the input elements were not stored; and c) in supplying connections across gaps in the physically specified input.

These criteria imply that abstract representations are *symbolic* representations. If a

representation encodes an object's or a contour's shape, it designates – symbolizes -- a property of a material object in the world (c.f., Newell & Simon, 1976). One of the deepest complexities of visual perception is that it involves a transition to symbolic descriptions of the environment from initially *subsymbolic* inputs. Early processing in the visual pathway involve units that respond to light or contrast. Encoding these properties of light energy, while crucial to vision, is not the *goal* of vision; rather, representing material properties of the world, such as objects, arrangements, and events are goals of vision (Gibson, 1966, 1979; Marr, 1982). The transition in visual perception from subsymbolic to symbolic coding appears to correspond to the distinction between coding of properties of incident light energy vs. representing properties of objects in the world. Arguably, the activation of a photoreceptor in the retina or a detector of oriented contrast in V1 are not symbolic, in that these activations do not represent properties of objects in the world. On the other hand, the perceptual description that one is seeing a rectangular table is symbolic. Representation of properties of objects, such as their unity, continuity, shape, and the material of which they are composed, are symbolic representations. A related notion of symbolic descriptions is that they include structured codes in which instances can be described by a relatively constrained set of numerical variables.

The transition from subsymbolic to symbolic processing remains deeply mysterious in visual perception (Kellman, Garrigan & Erlikhman, 2013), and most research occurs on one side or the other of this divide. Although the research reported here focuses specifically on understanding the representation of contour shape, it also has a larger purpose of using this domain as an example and existence proof of how the visual system may obtain, symbolic descriptions from initial subsymbolic encodings.

Consideration of how the visual system makes the transition from subsymbolic to symbolic representations led us to propose that 2D contours are organized into regions of similar curvature and recoded as a set of constant curvature primitives. There is evidence from research on early cortical areas that the visual system is sensitive to oriented luminance contrasts of different sizes across visual field (Hubel & Wiesel, 1962). Under the constant curvature theory, these local oriented contrast detectors are connected to larger curvature operators. These larger operators signal a unique curvature value but have tolerance for some variation in the true turn angle between local oriented contrast detectors. They are therefore abstract in the sense that they compress many similar curvatures into a single curvature value and symbolic in that the physical features of an object's contour are represented by a small number of descriptive variables.

Some theories of shape fail to capture the abstract qualities of shape representation. For example, the perceptual symbol system (Barsalou, 1999; Barsalou, 2003) proposes that we simulate activations from early subsymbolic areas when accessing a representation of an object. Deep convolutional neural networks also appear to never make the transition from subsymbolic activations from local elements to representations of the relations of parts (Baker, Lu, Erlikhman & Kellman, 2018; Brendel & Bethge, 2019; Baker, Lu, Erlikhman & Kellman, 2020). Many other theories of shape representation are abstract, but it is unclear how they could be derived from subsymbolic inputs. For example, maximum *a posteriori* skeletal representations of shape (Feldman & Singh, 2006) abstract the contour in impressive ways, but are computed only through complex probabilistic algorithms that are divorced from subsymbolic inputs in early vision. Many computer vision algorithms use sophisticated mathematical tools to capture shape representation in, for example, flux graphs (Rezanejad & Siddiqi, 2013) or quadric polynomials (Bolle & Cooper, 1984). As engineering solutions, these systems might work well, but no

explanation is given for how they connect to outputs of subsymbolic systems. It has been argued that these models operate at the computational level of description and that the elaborate neural circuitry at the brain's disposal can implement any good solution, but it also seems sensible that models in which the neural circuitry is plausible and straightforward are preferable.

The theory of constant curvature has been supported by several sources of empirical evidence. Contours already made up of constant curvature were found to take less time to encode than contours made up of continuously changing curvature (Kellman & Garrigan, 2011). Constant curvature contours are also easier to detect among random distractors and to segment from each other than contours that differ in other ways (Baker, Garrigan & Kellman, 2020). In this dissertation, I build on the theory of constant curvature shape representation and test the hypothesis that constant curvature representations are stable and give predictable outputs that do not depend on task demands, individual differences, or visual attention. This means that we can develop models of shape representation that make specific predictions about perceptual performance. I also use the idea of constant curvature representations to clarify certain perceptual phenomena, such as why some dot arrays are seen as a shape while others are not and why two shapes with different boundary elements are perceived as the same.

Ideas about abstraction in shape representation and the possibility that representations are built with segments of constant curvature led us to explore the phenomenon of shape representations forming from some configurations of unconnected dots. Unlike in complete objects, where the physical shape is specified by the object's edge points, dot arrays do not have any explicit boundaries or structure. Mathematically, the visual system could interpolate contours between dots in an infinite number of ways or interpolate no contours at all. Still, experimental evidence suggests that we construct stable representations of shape from certain

arrangements of dots (Baker & Kellman, 2018). Understanding what spatial relationships between dots give rise to a shape percept gives critical insight into what kinds of computations the visual system makes in the formation of an abstract shape representation. We studied the percept of curvature in dot arrays, testing a prediction from the constant curvature theory that shape representations would be more easily formed by arrays that can be encoded with relatively few curved segments than by straight segments joined at corners.

The transition from subsymbolic to symbolic shape representations also led us to consider objects that have a great deal of high frequency contour information along their boundary. In early visual areas, local edge detectors give dissimilar outputs for two objects that have the same global form but different local features. Still, when we see a cloud shaped like a dog or two pine trees with different local element features, the visual system finds a match between locally different objects with similar overall form. This equivalence can only be obtained at a symbolic level of contour description. The separation of global features from local elements also connects with the constant curvature theory of shape description, as one way of explaining how similar global forms are obtained from dissimilar local elements involves finding regions of similar curvature from large-scale curvature detectors proposed in the constant curvature model. Other possible explanations require interactions between local and global processing systems that experimental data do not support.

This document consists of four unpublished manuscripts, all of which pertain to abstract shape representation and constant curvature coding. They are written to be read independently but all touch on similar themes. In Chapter 1, I discuss a model for how a shape could be organized into relatively few segments of constant curvature at the computational level of description. I report psychophysical experiments that independently estimate the model's two

free parameters, then test whether outputs of the computational model are predictive of human performance on a shape matching task. In Chapter 2, the model is reformulated to operate at the algorithmic level of description, starting from outputs of early detectors and forming a symbolic description of an object's shape. In Chapter 3, I look at the perception of curved contours in dot arrays and analyze how experimental findings align with curve-based theories of shape representation. In Chapter 4, I propose a theory about how the visual system handles local and global information. I present empirical evidence that local and global contour features are processed in independent systems and that the local system primarily encodes statistical information about feature properties.

Taken together, the results of these studies support the idea that shape is abstract and separate from its local elements. They also support the notion that shapes are represented with segments of constant curvature in the visual brain and give a way that these representations could form from subsymbolic activations of early visual areas. These connections to early visual areas suggest solutions to several phenomena observed in reported experiments, such as how dot arrangements and noisy contours can be abstracted over to encode global form.

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, *61*(3), 183.

Baker, N., & Kellman, P. J. (2018). Abstract shape representation in human visual perception. *Journal of Experimental Psychology: General*, *147*(9), 1295.

Baker, N., Garrigan, P., & Kellman, P.J. (2020). Constant Curvature Segments as Building Blocks for 2D Shape Representation. In revision for *Journal of Experimental Psychology: General*.

Barsalou, L. (2003). Situated simulation in the human conceptual system. *Language and cognitive processes*, *18*(5-6), 513-562.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences*, *22*(4), 577-660.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. *Grounding cognition: The role of perception and action in memory, language, and thought*, 129-163.

Bolle, R. M., & Cooper, D. B. (1984). Bayesian recognition of local 3-D shape by approximating image intensity functions with quadric polynomials. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4), 418-429.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014-18019.

Gibson, J. J. (1966). The senses considered as perceptual systems.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston, MA, US.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106-154.

Kellman, Philip J., Patrick Garrigan, and Gennady Erlikhman. "Challenges in understanding visual shape perception and representation: Bridging subsymbolic and symbolic coding." In *Shape Perception in Human and Computer Vision*, pp. 249-274. Springer, London, 2013.

Marcus, G. (2001). The algebraic mind.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information.

Newell, A., & Simon, H. A. (2007). Computer science as empirical inquiry: Symbols and search. In *ACM Turing award lectures* (p. 1975).

Overlan, M.C., Jacobs, R.A., Piantadosi, S.T. (2017). Learning abstract visual concepts via probabilistic program induction in a language of thought, Cognition, Volume 168, pp. 320-334.

Rezanejad, M., & Siddiqi, K. (2013). Flux graphs for 2D shape analysis. In *Shape perception in human and computer vision* (pp. 41-54). Springer, London.

Constant Curvature Modeling of Abstract Shape Representation

Nicholas Baker[1*], Philip Kellman[1]

1. Department of Psychology, University of California, Los Angeles, Los Angeles, California, United States of America

* Corresponding Author

Email: nbaker9@ucla.edu (NB)

Abstract

Abstract shape representation is an important unsolved problem in human visual perception. Previous work has suggested that the visual system encodes shapes with a set of connected constant curvature primitives. In this study, we describe a model for how the visual system might recode a set of boundary points into a constant curvature representation. The model includes two free parameters that correspond to the degree to which the visual system encodes shapes with high fidelity vs. the importance of simplicity in a shape representation. In Experiment 1, we tested the visual system's sensitivity to differences in curvature. Subjects had to discriminate a contour made up of two curvature segments from one made up of one curvature segment. Subjects reached a 75% threshold when the ratio of curvatures between the two segments was 1.18:1. In Experiment 2, we tested the amount of curvature variation over which the visual system abstracts. We measured subjects' ability to discriminate a contour generated from cubic splines from a constant curvature approximation of the contour, generated at various levels of precision. Results indicated that up to a certain level of precision, subjects could not detect a difference between the original contour and a constant curvature approximation thereof, but there is a clear transition point beyond which detection performance monotonically improves. The results of these two experiments were used to fix the free parameters in our constant curvature model. We then assessed the validity of the parameterized model by generating shape pairs that were matched in physical similarity but differed in qualitative properties of their proposed constant curvature representations. These qualitative properties were found to be predictive of human performance in a shape recognition task. The results of this study provide evidence for the use of constant curvature shape representation in human visual perception and provide a testable model for how abstract shape descriptions might be encoded.

Shape is a fundamental aspect of object perception and recognition (Palmer, 1999; Biederman & Ju, 1988; Elder & Velisavljević, 2009; Landau, Smith & Jones, 1988; Imai, Gentner, Uchida, 1994; Xu, Carey & Quint, 2004). While shape can be defined in a physical sense as the set of points or curvatures along the bounding contour of an object, research in perception has found considerable evidence that shape representations differ from the physical features of a contour (Koffka, 1935; Pomerantz, Sager & Stoever, 1977; Kovacs & Julesz, 1993; Altmann, Bulthoff & Kourtzi, 2003; Kanizsa, 1976). Human visual shape representations appear to be a symbolic encoding of the shape contour which requires meaningful processing time to compute (Kellman, Garrigan & Erlikhman, 2013; Baker & Kellman, 2018).

One of the major reasons shape representations require abstraction is that they must be recognizable across a variety of viewing conditions (Rock, 1977). Research has shown that shapes and objects can readily be recognized across transformations in position (Biederman & Cooper, 1991; Lueschow, Miller & Desimone, 1994), size (Biederman & Cooper, 1992; Cooper, Schacter, Ballesteros & Moore, 1992; Ito, Tamura, Fujita & Tanaka, 1995), and orientation, both within the picture plane (Schmidt, Sprote & Fleming, 2016; Baker & Kellman, 2018) and in depth (Pizlo & Stevenson, 1999). Constancy across proximal changes to an object's projection onto retinae requires abstract recoding of the stimulus from a pattern of luminance contrasts to a symbolic, object-centric description of the spatial relationships between contour features in an object. This recoding requires time for the visual system to compute. Psychophysical evidence shows that comparisons between transformed shapes are only possible after processing time beyond what is needed for registration of the position of local elements (Baker & Kellman, 2018) or the formation of a visual icon (Bertamini, Palumbo & Redies, 2019).

Abstraction is also important for reducing the total amount of information along an object's contour. While a huge amount of unprocessed information is available immediately after sensation (Sperling, 1960; Cotlheart, 1980), a much smaller amount is encoded in more durable memory stores (Luck & Vogel, 1997). Capacity for shape information may not be as limited as for conjunctions of visual features, but experiments testing detection of contour differences suggests that some contour features are not encoded in our shape representation (Barenholtz, Cohen, Feldman & Singh, 2003). As another example, consider Figure 1. Although the two shapes have none of the same curvature values, they are visually indistinguishable. Attneave (1954) proposed that the visual system preferentially encodes regions of high curvature along a contour, arguing that, from an information theory perspective, high curvature areas are more informative about the object's shape (see also Feldman & Singh, 2005).

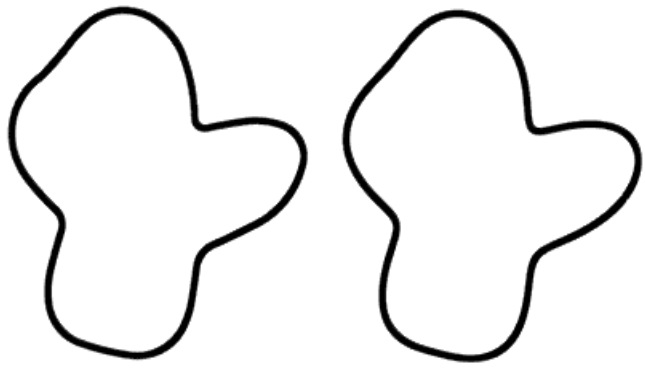

Figure 1. Two shape contours. The two shapes differ in local curvatures but give rise to the same shape percept.

Many models of shape representation have been proposed, both from work in human perception and computer vision. One of the most prominent theories is that shapes are encoded as a set of skeletal branches that capture areas of local symmetry along a contour. This idea originates with Blum (1973), who proposed the medial axis transform (MAT), which recodes a set of contour points to a set of axial branches. The MAT is completely data-driven and gives a

perfect reconstruction of the original contour. It is therefore abstract in the sense that it can capture potentially important perceptual features of a contour, but it does not simplify the contour representation, nor would it be robust to small, imperceptible changes to an object's contour. Newer skeletal models have corrected this, putting contour reconstruction accuracy in tension with representational simplicity (Feldman & Singh, 2006). These models are theoretically rigorous, but experimental evidence of their use has been limited. A few recent studies have found evidence that subjects' similarity judgments for objects correlate with the objects' skeletal similarity (Briscoe, 2008; Lowet, Firestone & Scholl, 2018; Ayzenberg & Lourenco, 2019). Crucially, however, the stimuli used in these experiments were generated from skeletal branches. Whether skeletal models are predictive of perceptual performance on shapes generated without a skeleton in mind remains an open question. One study found that there was no difference in subjects' ability to do a same/different task for shapes that were metrically different (i.e., a difference in the number of branches) vs. shapes that were qualitatively different (i.e., a change in the curvature along a branch) (Baker & Kellman, 2017).

Other, contour-based models of shape representation have also been put forward. Though not a fully specified model, Hoffman & Richards (1984) proposed a theory for how a shape is decomposed into parts by identifying concavities along the contour. Subsequent research has found support for the notion that curvature minima are more salient than curvature maxima along a contour (Barenholtz et al., 2003). Kass, Witkin, and Terzopolous (1987), modeled shape representations as a series of deformations of a basic shape primitive. They used evolving splines to capture these deformations. Several newer models have used the same idea of template deformation but adjusted the algorithms to make the shape representations more robust to local

contour changes and partial occlusion (Kimia Tannebaum Zucker, 1995; Elder, Oleskiw, Yakubovich & Peyré, 2013).

Another model of shape that has been used in computer vision research involves dividing a contour into a set of line segments, then merging together adjacent segments that are sufficiently similar in orientation (Pavlidis, 1982). Gdalyahu and Weinshall (1999) showed how split-and-merge techniques with line segment primitives could be used as the basis for object recognition algorithms in natural images. Similar split-and-merge methods have been proposed with splines rather than straight segments to more efficiently code curvature along a contour (Lancaster & Salkauskas, 1990). Another possibility is to decompose shape contours into segments of constant curvature. Wuescher and Boyer (1991) developed an algorithm for constant curvature segmentation of contours and showed that they fit contours better and with fewer primitives than straight segment approximations.

The idea that 2D shape representations are made up of smoothly joined constant curvature segments is intriguing for several reasons. From a neurophysiological standpoint, Pasupathy and Connor (2001; 2002) found evidence of neuron populations in V4 of neurons tuned to a specific curvature and position. They argued that representations of part-based shape could be built up from populations of these neurons. From an ecological perspective, a great deal of work in natural scene statistics has examined the prevalence of co-circular contours in our visual environment (Sigman, Cecchi, Gilbert & Magnasco, 2001; Chow, Jin & Treves, 2002). While results differ on how many truly co-circular contours exist in the visual environment, there is agreement on the prevalence of curved, nearly co-circular contours that could be well-estimated by constant curvature encoding.

Garrigan & Kellman (2011) also found empirical evidence for the use of constant curvature primitives in contour representation. Open contour fragments made up of constant curvature segments were encoded more accurately under brief viewing durations than fragments made up of non-constant curvature. They attributed this to the greater similarity between the physical properties of the constant curvature contour and subjects' abstract representation of the contour. Further tests have found evidence for the use of constant curvature primitives in contour representation, showing that constant curvature paths are easier to detect in visual search than non-constant curvature paths, and that people can learn to segment a contour made of two constant curvature segments much more accurately than a contour made of two segments with different physical properties (Baker, Garrigan & Kellman, under review).

In the current work, we adopt the hypothesis that abstract shape representations are built up from constant curvature primitives. Garrigan (2006) proposed a computational model for how a 2D contour might be represented a set of smoothly joined constant curvature segments. We briefly review the model below.

Model

The constant curvature model takes an object's bounding 2D contour as an input and outputs a representation of the shape made up of a small number of constant curvature primitives.

The model begins by computing the signed curvature at every point along the inputted shape contour. Signed curvature k is calculated as $\frac{\mathrm{x}'\mathrm{y}''-\mathrm{y}'\mathrm{x}''}{(x'^2y'^2)^{\frac{3}{2}}}$.

Next, the contour is segmented into regions of similar curvature by identifying points at which the curvature changes from higher than the local average to lower than the local average, or vice versa. The segmentation process considers all adjacent points, *a* and *b*, along a contour

and places a segmentation boundary between them if the difference between the curvature at *a* and the mean curvature in a local window centered on *a* is positive and the curvature at *b* and the mean curvature in a local window centered on *b* is negative, or vice versa. The precision of this segmentation depends on the size of the local window, which we term the integration window, *W*, with which the curvatures at *a* and *b* are compared. Formally, if

$$(\mathrm{k_a} - \frac{1}{2W+1}\sum_{i=a-W}^{i=a+W} k_\mathrm{i})^* (\mathrm{k_b} - \frac{1}{2W+1}\sum_{i=b-W}^{i=b+W} k_\mathrm{i}) < 0$$

then the model will add a segment boundary between *a* and *b*. Here, *W* represents the amount of contour considered when deciding if a segment boundary exists between *a* and *b*. Mean curvatures are calculated in the interval (*a* – *W*, a + *W*) and along the interval (*b* – *W*, b + *W*). Larger values of *W* correspond to larger windows that are therefore more tolerant to variance in curvature when the model decides whether to partition the contour between points *a* and *b*. Smaller values correspond a smaller window that is less tolerant to curvature variation.

Once the segment boundaries have been identified, the model recodes all contour points between the segment boundaries into a single constant curvature segment. A constant curvature segment is a contour region in which all points within the region are represented with the same curvature. It is described by an object-centric spatial position, a signed curvature (defined as the mean of all curvatures between the segment boundaries), and an arclength.

Curvatures of adjacent segments are then compared to ensure that all segments are sufficiently dissimilar to merit a separate primitive representation. If the difference in curvature between any adjoined segments is below a threshold value, *T*, then the visual system will represent the pair of segments a single primitive. The curvature of the merged segment will be the mean of the curvature of the two segments, weighted by their respective arclengths. This

process continues until no pair of adjacent segments have a curvature difference below *T*. In mathematical terms, if

$$|k_2 - k_1| < T$$

where k_1 and k_2 are the curvatures of adjacent segments, then the visual system will merge them into a single constant curvature segment with the weighted mean of their curvatures, given by

$$k_{12} = \frac{l_1 * k_1 + l_2 * k_2}{l_1 + l_2}$$

where l_1 and l_2 are the lengths of the adjacent segments. Here, *T* specifies the visual system's sensitivity to differences in curvature between constant curvature segments.

The final representation is the set of constant curvature segments composing the shape, each described by a position, curvature, and arclength (see Garrigan, 2006) for a more detailed description of the model).

The constant curvature model includes two free parameters: the size of the integration window (*W*) used in segmentation and the minimum difference in curvature (*T*) needed for two adjacent segments to be represented separately. Both of these parameters balance the representation's fidelity to the original contour with a preference for simpler representations built up from fewer primitives. This tension is a classic concern in research on visual perception (von Helmholtz, 1909/1962; Hochberg & McAlister, 1953; van der Helm, 2000). Research on *minimum tendencies* has found that the visual system tends to represent visual information as simply as possible, up to a certain loss in fidelity to the original stimulus (Hochberg, 1964; Buffart, Leeuwenberg & Restle, 1983; Hatfield & Epstein, 1985).

In shape perception, this tension has been formulated in Bayesian terms as the balance between a simplicity prior, where, in the absence of data, a representation that is more complex has less *a priori* probability, and a likelihood, where a representation is more probable if it more

closely matches the original contour (Feldman & Singh, 2006). The best representation, then, is one that matches the original contour reasonably well, but does not represent contour features that greatly increase the representational complexity while only marginally improving the representational fidelity. Note that the prior specified in Feldman and Singh's model has no relationship with frequency statistics in typical visual environments. Their Bayesian formulation is mathematically equivalent to non-probabilistic models that define cost functions for both complexity and differences from the true contour. Of course, a Bayesian framework still has flexibility in how the prior and likelihood are quantified. A model with a narrow prior distribution and a wide likelihood distribution will emphasize simplicity, while one with a wide prior distribution and a narrow likelihood distribution will emphasize fidelity.

In the constant curvature model, fidelity and simplicity are balanced by the values fixed to the integration window size (*W*)and curvature difference (*T*) parameters. For *W,* using a smaller window results in more segment boundaries approximating the original contour. Representations formed from small window sizes therefore have higher fidelity but more complex than representations formed from larger window sizes. For *T*, a larger threshold results in the combination of segments with larger curvature differences, resulting in fewer primitives but greater difference between the physical and represented curvature in a region of the contour.

In order to test specific predictions of the constant curvature model, both of its free parameters must be specified. It is an open question how consistent the parameters are across people and viewing conditions. Garrigan (2006) hypothesized that they are flexible, and that the visual system uses smaller parameters for visual tasks that require a high degree of specificity, and larger parameters for tasks in which a loose approximation of the shape is adequate. On the other hand, if parameters are truly believed to vary flexibly, the visual system would be unable to

match same shapes that were viewed under different task conditions. It is possible that additional perceptual resources could be allocated for tasks that require an extremely high degree of shape fidelity, but it seems possible that the visual system always encodes a representation with a fixed level of specificity for general recognition.

In the current work, we tried to estimate and evaluate the constant curvature model specifically as used for shape recognition. In Experiments 1 and 2, we used psychophysical experiments with simple open contour stimuli to fix the curvature difference threshold (Experiment 1) and integration window size (Experiment 2) in the computational model for constant curvature representation. In Experiment 3, we used the parameters fixed in Experiments 1 and 2 to test the empirical validity of the fully parameterized constant curvature model.

Experiment 1

In Experiment 1, we aimed to specify the threshold parameter (T) of the constant curvature model. The threshold parameter determines the point at which two adjacent constant curvature segments are represented as a single segment based on the difference in curvature between them. In other words, if the curvature difference between two smoothly connected segments is less than the threshold parameter, then the visual system should encode them as a single segment of constant curvature that extends the length of both constituent segments. The parameter acts in service of the simplicity constraint for the constant curvature model, ensuring that a shape is represented by as few primitives as possible provided that the representation still has sufficient descriptive capability to support recognition, discrimination, motor action, and reasoning about functional properties.

We used a simple psychophysics experiment to measure the visual system's sensitivity to differences in curvature between two smoothly joined constant curvature fragments. Previous work by Baker et al. (2020, under review) found that the visual system is capable of accurately

separating two constant curvature segments at their point of transition, but not two segments of constantly accelerating curvature. In this experiment, we varied the difference in curvature between two CC segments in order to determine at which curvature differences the two segments are perceived as a single curvature as opposed to a composition of two distinct curvatures. We hypothesized that the maximum curvature difference that was undetectable to participants would be a natural threshold for the constant curvature model in deciding whether two adjacent segments should be represented as one or two segments of constant curvature.

Method

Participants

Twenty-six undergraduates (eight male, 18 female, M_{age} = 20.6) from the University of California, Los Angeles participated in Experiment 1 for course credit. All subjects had normal or corrected-to-normal vision.

Stimuli

In each trial, we generated two open contour stimuli, one made up of a single constant curvature segment, and one made up of two smoothly joined constant curvature segments. The single CC contour was generated with a random length and curvature. The length of the contour was between 240 and 500 pixels (5.76-12 degrees of visual angle), and the curvature was between 0.0059 and 0.020 pixels^{-1}. Both contours made up of a single CC segment and contours made up of two constant curvature segments had lengths between 120 and 250 pixels. The mean length of contours for both conditions was equated. The angular extent of the single CC contour was determined by its length and curvature, but we added a constraint that the angular extent must be less than 360 degrees.

The contour made of two connected CC segments was created by generating one constant curvature segment with random length (between 120 and 250 pixels) and curvature (between 0.0059 and 0.020 pixels^{-1}), then smoothly connecting a second segment to it. For the contour to be differentiable at all points, the second segment had to begin at the same angular position at which the first segment terminated. The second segment also had random length between 120 and 250 pixels. The curvature of the second segment was determined by the first. We varied the difference in curvature across nine conditions as a ratio. The possible curvature ratios ranged from 1.03:1 to 1.9:1. The order of the two segments was randomized so that half the time the higher curvature segment was clockwise of the lower curvature segment, and the other half it was counterclockwise.

Display and Apparatus

Subjects were seated 70 cm from a 20-in. View Sonic Graphic Series G225f monitor. The monitor was set to 1024 x 768 resolution, with a refresh rate of 100 Hz. All stimuli were black contours shown on a gray background. One contour was shown in the center of the left half of the screen, and the other was shown in the center of the right half of the screen.

Design

Experiment 1 had 225 trials, consisting of nine conditions with 25 trials each. The nine conditions corresponded to the ratio of curvatures between the two constant curvature segments from the two CC open contour. The nine ratios were 1.03:1, 1.06:1, 1.1:1, 1.3:1, 1.4:1, 1.6:1, 1.75:1, and 1.9:1. The conditions were ordered in blocks from the highest ratio to the lowest so that better performance for higher ratios could not be explained by practice effects. Subjects completed five practice trials with the experimenter present to ensure that they understood the instructions before beginning the main experiment.

Procedure

In each trial, one open contour with a single curvature segment and one open contour made of two curvature segments were shown simultaneously, one in the left half of the screen and one in the right half. Position was randomly assigned in each trial. Participants were told to look at both open contours and judge which one they believed was more complex. They were then told to press "A" if they believed the contour on the left was more complex, or "L" if they believed the contour on the right was more complex. No explanation was given about what was meant by complexity, but the more complex stimulus (i.e., the open contour made of two constant curvature segments) was highlighted in blue in each trial after subjects had given their response. Subjects were free to look at the stimulus pair for as long as they wanted before responding.

Results

Mean accuracy results are plotted for each curvature difference in Figure 2. Performance was at chance for a ratio of 1.03:1 ($t(25) = 0.18$, $p = .83$, 95% CI = [.47, .54]). At a ratio of 1.06:1, performance was marginally better than chance after correcting for multiple comparisons ($t(25) = 3.71$, $p = .01$, 95% CI for performance = [.53, .62]). For all other curvature ratios, subjects performed reliably better than chance even with a Bonferroni correction for multiple comparisons.

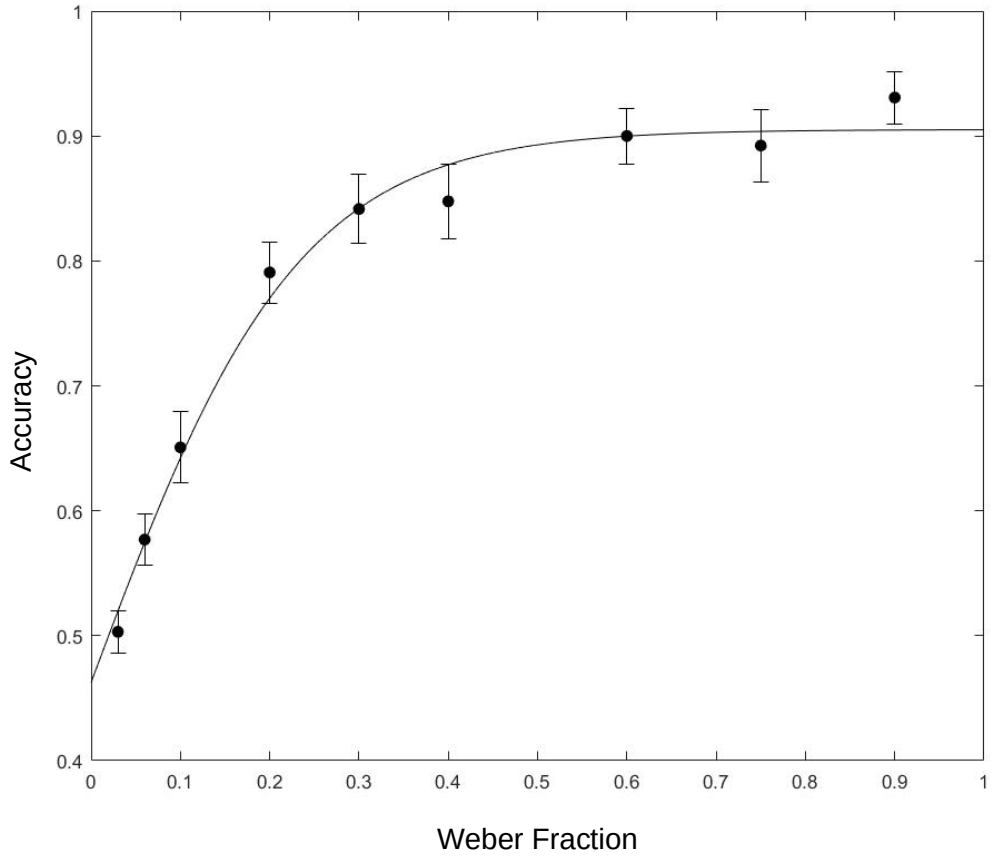

Figure 2. Performance as a function of curvature difference. The horizontal axis gives the Weber fraction between curvatures for the two-segment contour fragment. For example, if the Weber fraction is 0.1, then the curvature ratio between the two segments is 1.1:1.

Performance improved rapidly with larger curvature ratios up to a ratio of 1.3:1, after which it flattened out and improved only marginally as the ratio got larger. This suggests that the critical point at which the visual system encodes two contours of similar curvature as different occurs somewhere between a ratio of 1.03:1 and 1.3:1. Since chance was 50%, we found the 75% performance threshold to estimate the value of *T*. We fit a psychometric function to the data using the Palamedes Toolbox (Prins & Kingdom, 2009) and found the 75% threshold to be a curvature ratio of 1.18:1.

Discussion

Much of the foundational work in psychophysics was aimed at identifying the point at which a physical difference in a stimulus is detected by sensory systems (Weber, 2018; Fechner, 2012). In Experiment 1, we used similar psychophysical methods to evaluate subjects' perceptual ability to detect differences in curvature. Unlike the oldest detection work, this study did not test subjects' sensitivity to energies such as light or sound, but to object features.

Despite a physical difference in curvature, the visual system appears not to be able to detect curvature differences between contours when the curvature ratio is very small. Participants did not reliably perceive the contour made of two constant curvature segments as more complex when the ratio was 1.03:1, and perceived contours with a curvature ratio of 1.06:1 only marginally better than chance, presumably because these stimuli were encoded as a single segment of constant curvature. As the ratio between curvatures got larger, the probability of subjects encoding the contour with two curvature values increased, resulting in more accurate selection of the more complex contour. Subjects appear to reach ceiling accuracy when the curvature ratio is about 1.3:1 and are not more likely to judge the two-curvature contour as more complex for higher ratios of curvature.

Importantly, subjects were instructed to choose the more complex stimulus, not the stimulus with more than one curvature. In fact, no mention of curvature was made at any point during Experiment 1. Still, for all but the most similar curvature pairs, subjects automatically chose the contour made of two curvature segments as the more complex stimulus. The ease with which subjects used curvature difference as an indicator of stimulus complexity furnishes additional evidence, beyond earlier work (e.g., Garrigan & Kellman, 2011; Baker et al., 2020, under review) evidence that constant curvature segments are indeed a basic unit of abstract shape representations. If a shape is represented by a set of primitives, it stands to reason that shapes

built up from more primitives are more perceptually complex than shapes built up from fewer primitives.

Other notions of contour complexity do not make the same prediction. Hoffman and Richards (1984) hypothesized that a shape is decomposed into parts based on the presence of curvature minima. By this definition, the contours in Experiment 1 all have the same number of parts, as the sign of curvature never changes for either stimulus category. Our data suggest that there are perceivable differences in shape complexity even when part numbers are the same. Another account of contour complexity posits that stimuli with higher curvature are more complex (Attneave, 1954; Feldman & Singh, 2005). In our study, the one-segment contour will have higher curvature than the two-segment contour 50% of the time, but subjects reliably chose the two-segment contour as more complex. One explanation for this is that the visual system does not have lowest surprisal when a contour continues straight in the tangent direction as has been suggested (Feldman & Singh, 2005), but when its curvature is most similar to the curvature of contour areas nearby it.

Experiment 1 helps us to fix the threshold parameter for the constant curvature model. The segment merging operation in our model is deployed after an initial segmentation of the contour into constant curvature segments has already been completed. It serves to prune the shape representation by encoding adjacent segments of similar curvature with a single CC segment. Two segments are merged into a single primitive if the difference in curvature between them is below a certain threshold. A likely candidate for what this threshold might be is the point at which two curvatures are detectably different more often than not. By fitting a psychometric function to the Experiment 1 data, we found the 75% threshold as an estimate of when curvature differences are reliably detected. This point corresponds to a curvature ratio of 1.18:1.

In Experiment 1, we used only segments of constant curvature. Certainly, out in the world, object contours have far more curvature variation (Chow, Jin & Treves, 2002). For our purposes, however, restricting our stimuli to one or two segments of constant curvature is more directly relevant to how the threshold parameter works in the constant curvature model. The threshold parameter merges segments of similar curvature after they have already been segmented and recoded into constant curvature primitives (see Model), so we restricted our stimuli to contours that were plausible inputs at that stage of the constant curvature model.

Experiment 2

In Experiment 2, we estimated the second free parameter of the constant curvature model: the size of the integration window used in segmenting a contour into constant curvature regions. In our model, the integration window size is parameterized as *W*, and corresponds to a percentage of the contour's total length. *W* determines how contour regions with non-constant curvature are approximated by a relatively small set of constant curvature segments. In the constant curvature model, the visual system assigns adjacent points along a contour, *a* and *b*, to different segments if the difference between the curvature at *a* and the mean of curvature within the integration window centered on *a* is positive and the difference between *b* and the mean of curvature within a window centered on *b* is negative, or vice versa. *W*, then, determines how finely or coarsely a contour is segmented into regions of constant curvature.

If the model uses large integration windows, a higher percentage of contour points will be shared in the integration window centered on *a* and the integration window centered on *b*, so a contour boundary will be drawn more rarely between two points. Consequently, model outputs with large *W* will tend to be coarser, abstracting over more curvature variety along an object's contour. On the other hand, models in which *W* is small yield outputs with more constant

curvature segments, which are consequently much more visually similar to the original object boundary (see Figure 3).

Figure 3. Two constant curvature representations of a shape contour. The original shape (left) is approximated with a small *W* (middle) and a large *W* (right). The smaller integration window has more segments and represents the contour more precisely.

In the extreme, the visual system would represent every unique curvature along an object's boundary with its own CC segment. This would give a perfect reproduction of the contour but would also likely tax the visual system far beyond the capacities of visual memory (Farah, Rochlin & Klein, 1994). More likely, the visual system abstracts over some curvature variation, but encodes a precise enough representation to allow it to discriminate between similar but nonidentical shapes, such as those in Figure 4.

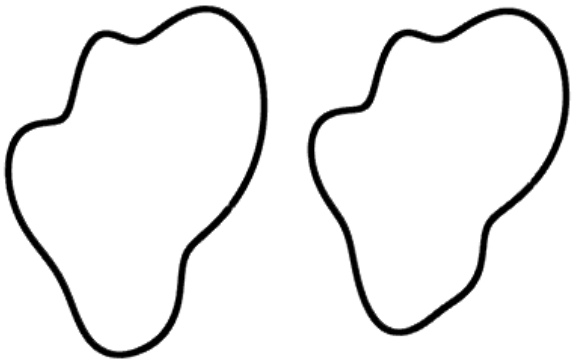

Figure 4. Shape pair with similar contour features.

The degree of precision with which contours are encoded is an empirical question. If the visual system is segmenting contours into constant curvature segments and encoding the constant curvature representation, there should be a point at which the constant curvature representation of a contour is indistinguishable from the original contour in a visual memory task. To determine this point, we compared subjects' ability to discriminate constant curvature representations of contour fragments from the real fragment across a variety of integration window sizes. We hypothesized that the visual system uses the largest integration window for which the model output indistinguishable from the inputted contour, since that will be the window size that is most economical while also representing the shape with sufficient precision.

Method

Participants

Twenty-three undergraduates (7 male, 16 female, M_{age} = 20.6) from the University of California, Los Angeles participated in Experiment 2 for course credit. All subjects had normal or corrected-to-normal vision.

Display and Apparatus

All display conditions were the same as in Experiment 1.

Design

Experiment 2 consisted of nine conditions, corresponding to nine integration window sizes. We initially specified window size as a contour length in degrees of visual angle. The nine sizes were 0.32, 0.64, 0.97, 1.29, 1.61, 1.93, 2.25, 2.58, and 2.90. There were 20 trials for each condition. In addition to the 180 experimental trials, we also had 180 trials (20 using each integration window size) in which the first contour fragment was made of constant curvature (see

Procedure). Results from these trials were not analyzed. Before beginning the experiment, subjects completed 10 practice trials.

Stimuli

Each trial included a contour fragment with nonconstant curvature. The contour was obtained by first generating a closed contour by displacing 12 control points along a circle and fitting cubic splines between the control points (see Figure 4 for two examples), then taking a fragment from the closed contour, totaling 40% of the closed shape's overall contour length, on average 12.88 degrees of visual angle. Every trial also had a constant curvature representation of the contour, generated with the fixed threshold parameter from Experiment 1 and various integration window sizes specified by the nine trial conditions.

Procedure

In the analyzed trials, subjects were shown a fixation cross in the center of the screen for 300 ms, followed by the nonconstant curvature contour fragment for 500 ms. A pattern mask was shown for 500 ms after exposure to the contour fragment, after which subjects were shown two contours simultaneously (one in the center of the left half the screen, one in the center of the right half of the screen) and asked which one exactly matched the first contour they had been shown. One of the two contours shown after masking was identical to the first contour. The other was the constant curvature representation of the contour generated with a window size determined by the trial condition (see Figure 5 for a sample trial). In the analyzed trials, the correct response was always the nonconstant curvature contour fragment. Subjects could view the two contours for as long as they wished before responding.

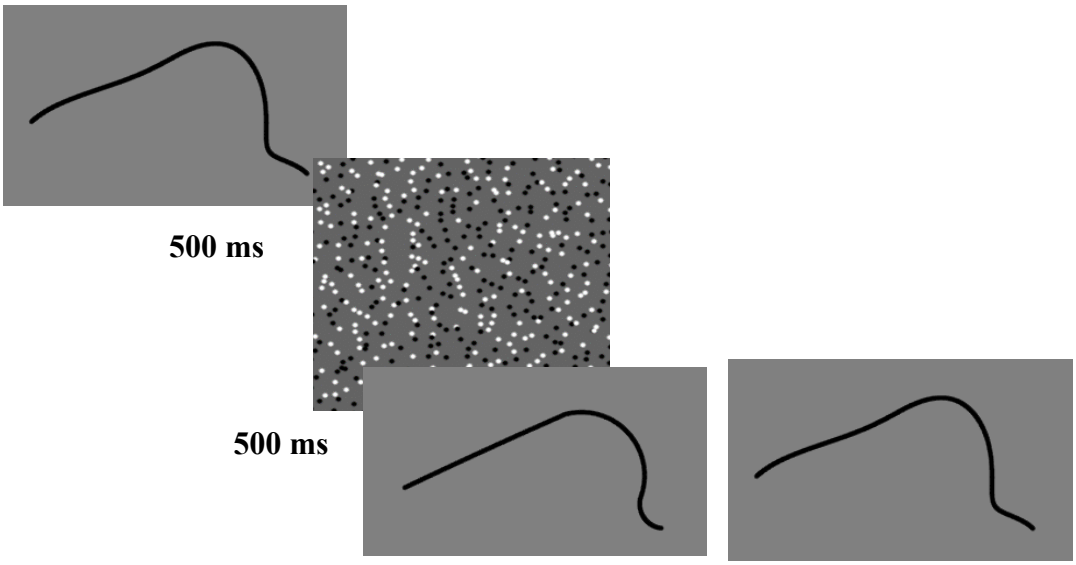

Figure 5. Sample trial for Experiment 2. The nonconstant curvature contour was shown first, followed by a mask. Then, a constant curvature representation and the original contour were shown side-by-side. Here, the constant curvature representation is generated with a window size of 2.25 degrees of visual angle.

To prevent subjects from using a strategy in which they always pick the contour fragment with constant curvature without comparison to the original contour, we also had 180 trials in which the constant curvature representation of the contour fragment was shown first instead of the nonconstant curvature contour fragment. After masking, the constant curvature contour and the nonconstant curvature contour were shown as in the main trials, but the correct response for the matching shape was the constant curvature contour. These trials were not analyzed because they do not require the visual system to abstract over local variations in contour curvature.

Results

The results of Experiment 2 are shown in Figure 6. After correcting for multiple comparisons, subjects are at chance performance for the three smallest integration window sizes ($t(21) = 1.75$, $p = .09$, 95% CI = [.49, .59], $t(21) = 2.25$, $p = .03$, 95% CI = [.51, .63], $t(21) = .91$, $p = .37$, 95% CI = [.47, .57], respectively). Starting at an integration window size of 1.29 degrees, subjects are able to reliably distinguish the constant curvature representation from the

original contour fragment ($t(21) = 3.77$, $p = .001$, 95% CI = [.53, .61]). Performance improves monotonically for larger integration window sizes.

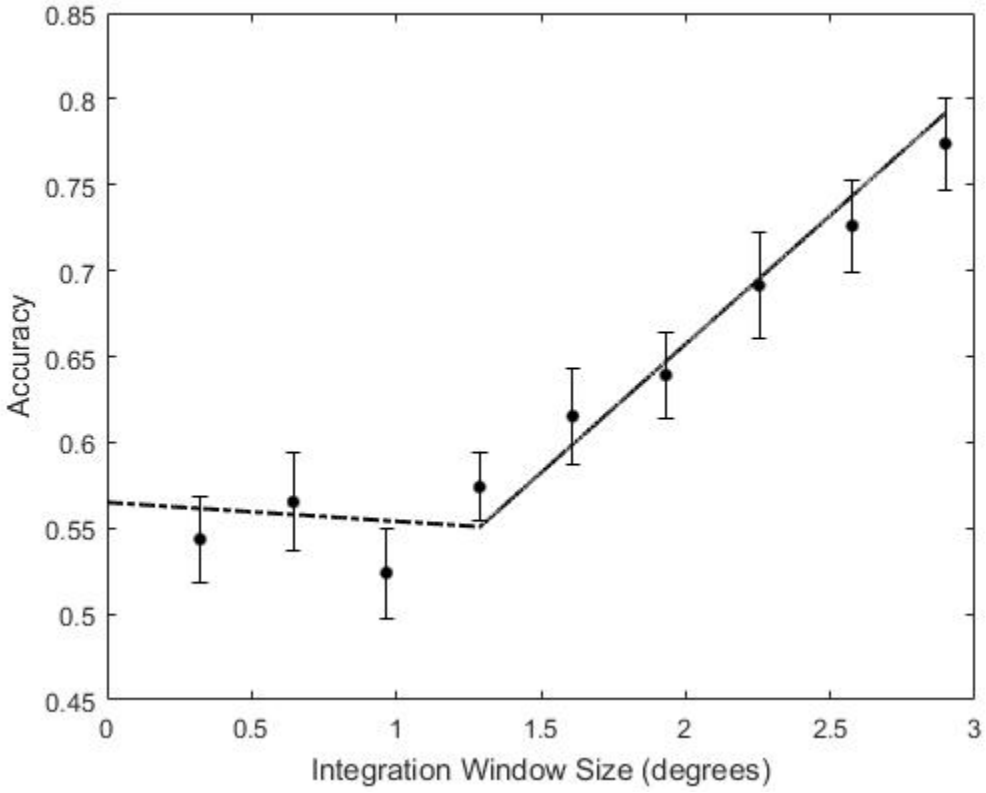

Figure 6. Results from Experiment 2 with best-fitting piecewise linear regression. The dashed line shows the linear fit up to the transition point and the solid line shows the linear fit past the transition point.

Unlike the data from Experiment 1, which followed an S-shaped function, the data in Experiment 2 appear to be well described by two linear functions, one approximately flat for integration windows that are all indistinguishable from chance guessing, and another function with positive slope beginning at the point where the constant curvature representation is coarser than the shape representation encoded in visual memory. Because we were looking for the very first point in which accuracy as a function of W is described by a positive slope, we analyzed the data for a change in slope polarity rather than looking for the 75% threshold as we did in Experiment 1. We looked for the smallest integration window size at which people begin to

detect that the constant curvature representation of the shape was different than the original shape. This corresponds to a transition point in a continuous piecewise linear regression model from zero (or, in our case, slightly negative) slope to a positive slope. To identify this transition point, we fit the data with several continuous piecewise linear regression models, specifying different sizes of W at which the slope changes in order to determine which one explained the most variance. R^2 was highest (.299) when the slope changed at W = 1.29 degrees of contour length, ($F(2, 204) = 43.52$, $p < .001$). For this regression, the slope before the transition point is not significantly different from zero ($t(2) = -0.69$, $p = .49$), while the slope beyond the transition point does significantly differ from zero ($t(2) = 2.69$, $p = .008$).

Discussion

In Experiment 2, we sought to fix the constant curvature model's second free parameter, the size of the integration window used to segment a contour into CC primitives. In the computational model, segmentation based on window size precedes CC segment merging governed by the threshold parameter we estimated in Experiment 1. The segmentation is an intermediate output in the model, and therefore not directly testable, so we estimated the threshold parameter first, then fixed it in the model for Experiment 2. Even though subjects were tested on model outputs with both parameters, only the integration window size was varied between trials, so performance differences can only be explained by the segmentation parameter.

Subjects had to encode an open contour into visual memory, then match that contour in a two alternative forced choice task. The distractor in the 2AFC task was always a constant curvature representation of the contour made with a varied integration window size and a fixed merging threshold parameter. When the integration window was small, subjects' encoded contour representation was indistinguishable from the constant curvature representation, despite

substantial physical differences between the original stimulus and the constant curvature stimulus. For example, the constant curvature representation generated with a 0.97-degree integration window could not be discriminated from the original shape contour at better than chance rate, despite having, on average, 3.7% as many unique curvatures (9.83 vs. 266.45).

What integration window size is most likely to be used by the visual system for constant curvature segmentation? Any window size that falls along the positive slope region of our regression is likely too large because subjects perceive a difference between the constant curvature contour and the representation they have encoded. On the other hand, if two model outputs are equally indistinguishable from the encoded contour representation, the visual system is most likely using the simpler model output for reasons of efficiency. The best choice for the integration window size, then, is the point of slope change in our continuous piecewise regression analysis, or 1.29 degrees of contour length.

In Experiment 2, we operationalized the integration window size as an absolute measure of length in terms of degrees of visual angle. This is likely not how it is used in the visual system. We would not, for example, expect a difference in segmentation for a shape viewed at different distances. There is a great deal of evidence that shape representations are scale invariant (Biederman & Cooper, 1991; Cooper & Schacter, 1992; Lueschow, Miller & Desimone, 1994; Ito, Tamura, Fujita & Tanaka, 1995). One possibility is that the size of the integration window is a percentage of the overall contour. The 1.29 degrees in our data would correspond to 10% of the contour's total length.

Issues arrive, however, when we use contour fragments as in this experiment. It seems unlikely that the visual system would segment a contour differently if it was fragmented from a larger contour or if part of the contour was covered by an occluder. Ideally, the integration

window size would be invariant to changes in scale, while still depending only on curvatures relatively nearby to it. One way we could correct for this is by using window sizes that are a percentage of the contour but adjusted by the sum of the turn angle within that contour. Since object contours tend to be closed, we would expect the whole contour to have a sum turn angle of 2π. We can therefore estimate the amount of the contour visible as a ratio of the sum of the turn angle for the visible contour to 2π. In Experiment 2, for example, the average turn angle for the open contours subjects were shown was 2.48 radians, or 39.5% of the whole contour's length. Then, instead of fixing the integration window size at 10% of the visible contour, we can fix it at 3.95% of the whole contour. We will use this formulation going forward because it allows us to make scale invariant segmentations that are nonetheless consistent across fragmentation and partial occlusion.

Experiment 3

Experiments 1 and 2 aimed to fix the two free parameters in the constant curvature model. In Experiment 3, we tested whether the parameterized constant curvature model explains human shape perception. Previous work by Baker et al. (2020, under review), found considerable evidence for the role of constant curvature segments as primitives of shape representations. However, it could not test predictions based on model outputs because the window size and curvature threshold were not fixed. Using the estimated parameters from Experiments 1 and 2, we tested whether features of the constant curvature representation of a shape pair could predict human performance on a matching task over and above physical contour differences.

Specifically, we tested whether differences in a shape that necessitate a new constant curvature segment are more detectable than shape differences over which the constant curvature model ultimately abstracts. We generated pairs of shapes by deforming the contour of a novel

shape by a small amount. We then subjected both members of the pair to the parameterized constant curvature model to determine whether the deformation necessitated fewer or more constant curvature segments in the representation. We tested subjects' sensitivity to a difference in shape for pairs that differed by zero, one, two, three, or four segments in the constant curvature representation produced by the model. If our model and parameters were correct, we predicted that shape pairs that differed by more constant curvature segments would be easier to discriminate from each other than shape pairs that differed by fewer segments, even if the physical difference between pairs was equated in both conditions. In general, a larger difference in segment number will correspond to a larger physical difference in two shapes' contours. To account for this, we fixed the physical difference in pairs of shape contours across conditions before testing whether two shapes that differed more in segment number were more perceptually different despite being equally physically different.

Methods

Participants

Twenty-three undergraduates (7 male, 16 female, M_{age} = 19.65) from the University of California, Los Angeles participated in Experiment 3 for course credit. All participants had normal or corrected-to-normal vision.

Display and Apparatus

All display conditions were the same as in Experiment 1.

Design

There were 250 trials in total. Experiment 3 consisted of six conditions which differed in the segment number difference between a pair of shapes. In the first condition (125 trials), the two shapes were identical. In the other five conditions (25 trials each), the shapes were different.

Conditions were separated based on the difference in number of segments between the two shapes in a pair, from zero to four. All conditions were interleaved with each other, and subjects were never informed of these different conditions. Before beginning the main experiment, subjects completed five practice trials.

Stimuli

Whole shapes were generated as in Experiment 2, by shifting 12 control points a random distance from a circle and radially fitting cubic splines between them. None of the contour regions in the shapes had constant curvature. "Different" shape pairs were generated by moving two adjacent control points a random distance to maintain equal contour length, and re-fitting a cubic spline between the new set of 12 control points.

For the different shapes, we wanted to create shape pairs that were equally different from each other across conditions. When we randomly generate shapes and subject them to the model, various displays will have different amounts of deviation in the representation from the physically given stimulus. They also may have different numbers of constant curvature segments. We wanted to create shape pairs that had equal physical similarity despite differing in number of segments by various amounts. In other words, if in the condition where one shape had one more constant curvature segment than its pair, the two members of the pair were 95% similar, then pairs of shapes in conditions in which the segment number difference was zero, two, three, or four should also have an average similarity of 95% To this end, we imposed a constraint on the physical difference between shape pairs, discarding pairs that were too physically different. We computed physical contour similarity by taking the ratio of the overlapping areas to the non-overlapping areas for both contours. Since this measure is asymmetrical, we computed the average:

$$\frac{\frac{\text{Shape 1 and 2 overlap}}{\text{Total area of Shape 1}} + \frac{\text{Shape 1 and 2 overalap}}{\text{Total area of Shape 2}}}{2}.$$

We generated hundreds of shape pairs and categorized them based on their constant curvature segment difference. We computed the segment difference by using the threshold parameter of 1.18:1 from Experiment 1 and the integration window size parameter of 4% of the whole contour from Experiment 2. After sorting them into five categories corresponding to a difference in number of segments of zero, one, two, three, or four, we confirmed that the physical difference of shape pairs was matched across categories. Mean difference and standard deviation for all five categories were matched to a hundredth of a percent at 97.51% similarity and 0.29% deviation. Sample pairs from each condition are shown in Figure 7.

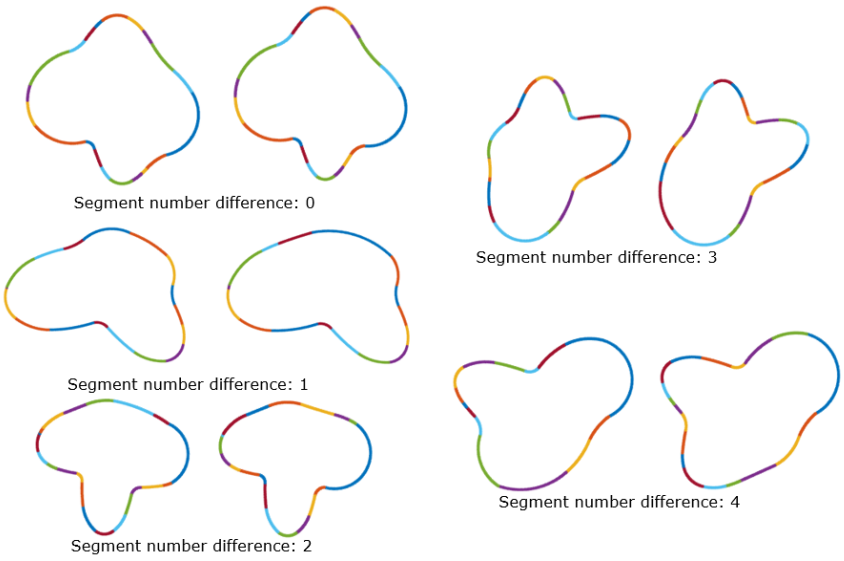

Figure 7. Sample shape pairs with different number of constant curvature segment differences.

Procedure

In different trials, participants were shown both members of a pair of shapes and were asked to judge if they were the same or different. Though equally physically similar, the pair of shapes could have a segment number difference between zero and four. In same trials,

participants were shown exactly the same shape twice. Each trial began with a fixation cross shown in the center of the screen for 500 ms, followed by the first shape, which remained on the screen for 250 ms. The first shape was then masked by a pattern of black and white dots (600 ms), after which the second shape was displayed, along with a prompt at the top of the screen asking subjects to decide if the second shape was the same or different as the first shape they had been shown. Participants were instructed to press A if the two shapes were the same, or L if the shapes were different. The second shape remained on the screen until subjects responded. Figure 8 shows a sample trial. Subjects completed five practice trials with the experimenter present to make sure all instructions were understood before beginning the main experiment.

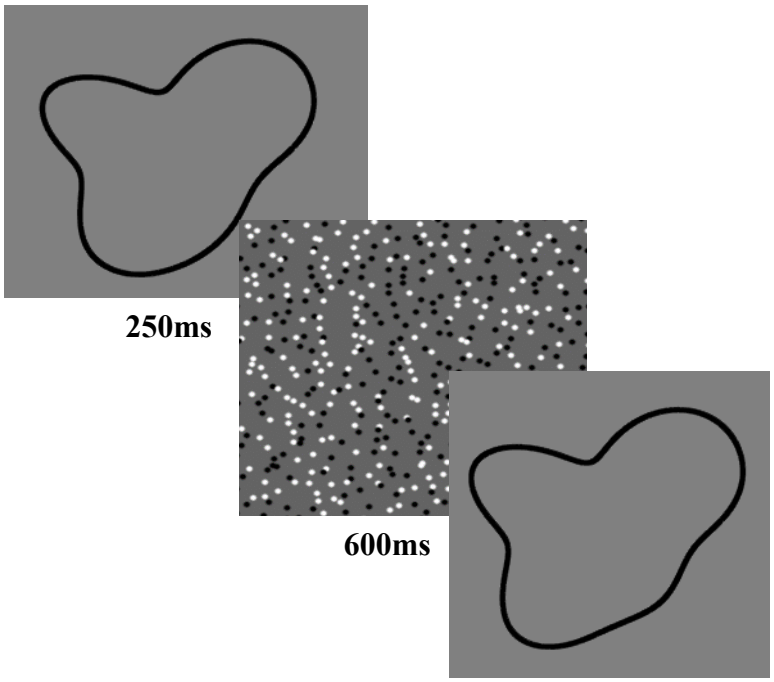

Figure 8. Sample trial from Experiment 3.

Results

One participant's data was not analyzed because she was at chance performance across all conditions. The overall results are not affected by her inclusion or exclusion. We analyzed the data from Experiment 3 by computing subjects' sensitivity to a shape change between the first

and second display. A hit was classified as a correct "different" response when the shape had changed, and a false alarm was classified as an incorrect "different" response when the second shape was the same as the first. The sensitivity results are shown in Figure 9.

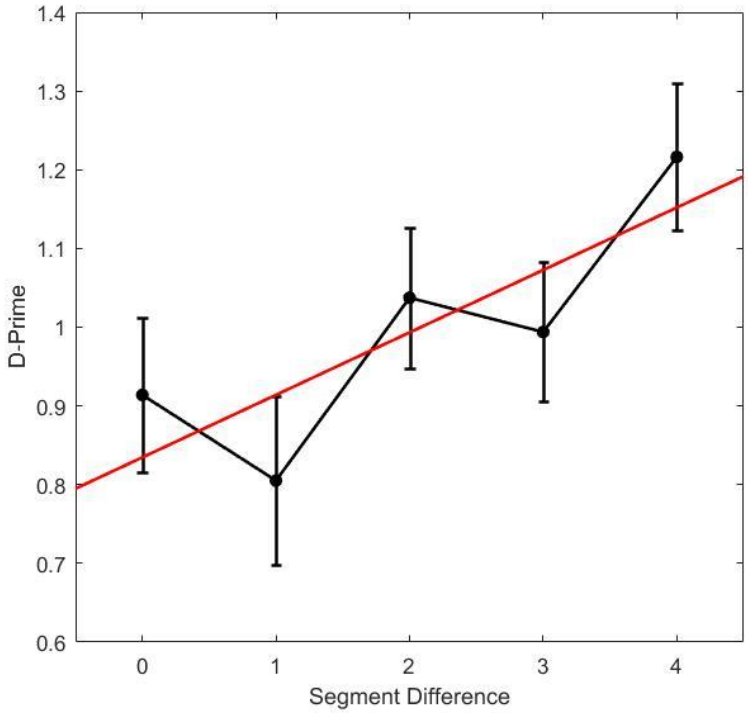

Figure 9. Results for Experiment 3.

To analyze whether there is a significant effect of segment number difference on subjects' ability to detect a shape change, we fit a linear regression function to our data. The data was best fit by the function $y = 0.83 + 0.08x$. A repeated measures ANOVA confirmed a significant linear component, $F(1,20) = 17.62$, $p < .001$, $\eta^2_{partial} = .468$, indicating a significant positive change in performance as the difference in segments gets larger.

Discussion

After fixing the constant curvature model's two free parameters, in Experiment 3, we aimed to assess the model by testing whether its outputs could explain aspects of human

perceptual capabilities. Experiment 2 furnished some evidence for constant curvature encoding of 2D shapes, as constant curvature representations generated with a sufficiently small segmenting window were indistinguishable from the original contour, which had no regions of constant curvature. In Experiment 3, we used one aspect of a constant curvature representation, the number of primitives from which it was composed, to assess the model's validity.

In our model, some differences in curvature are abstracted over as shapes are recoded into constant curvature regions, while other curvature differences necessitate a segment boundary. We predicted that those differences that result in a change to the number of constant curvature segments should be more salient to viewers. Consequently, when we generated shape pairs by deforming a contour, we predicted that if the contour changed did not lead to a segment number difference in the two shapes' constant curvature representation, a difference between the two shapes should be harder to detect than in a pair of shapes where the deformation resulted in the gain or loss of constant curvature segments. Moreover, we predicted that larger changes in segment numbers should make shape differences more detectable.

To test this, participants were shown shape pairs in sequence and told to do a same/different task. Participants were never told anything about the constant curvature model or given any kind of indication that the "different" shape pairs fell into five distinct categories. Nevertheless, they showed a clear difference in performance across the experimental conditions, getting better at the recognition task as the difference in constant curvature segments got larger. Importantly, differences in constant curvature number did not correlate with the magnitude of physical contour differences in this experiment. The distribution of contour differences was identical across all five conditions. Performance differences are therefore explained by how

shapes are perceived and encoded, not by differences in the contour that are independent of visual processing.

The results of Experiment 3 are congruent with the parameters we fixed in Experiments 1 and 2. Both the segmentation window and curvature threshold parameter balance representational efficiency with representational accuracy. In Experiment 3, if the outputted representations were too simple, contour differences that are important to our shape representations would not be captured and shape pairs that are perceptually different would often have the same number of curvature segments. If the representations were too precise, the model outputs would not sufficiently diverge from physical contour differences, making it difficult to predict performance differences between shape pairs whose physical contour similarities are equated.

General Discussion

In this work, we aimed to develop a model for abstract representations of 2D shape contours. Specifically, we focused on a curve-based model of shape representation in which contour regions are recoded as constant curvature segments. Constant curvature models of shape are appealing for neurophysiological (Pasupathy & Connor, 2002), ecological (Sigman et al., 2001; Chow et al., 2002; Wuescher & Boyer, 1991), and behavioral reasons (Wu, Frund & Elder, 2016; Baker et al., 2020, under review). In the current work, we went beyond testing whether constant curvature segments were plausible primitives for shape representation to propose and assess a specific model of shape encoding.

The constant curvature model is generative, taking physical contours as inputs and proposing an internal representation of the shape as it is encoded in visual perception. In biological vision systems, the transformation from a physical contour to an abstract

representation involves recoding initial subsymbolic representations (e.g., detection of oriented contrast) into symbolic representations. Subsymbolic representations respond monotonically to luminance energy in a visual scene and can be thought of as a literal record of the pattern of light striking the retina. While rich in information, these representations are inflexible to changes in viewing condition and brittle, easily destroyed by masking or the passage of time (Coltheart, 1980; Sligte, Scholte, & Lamme, 2008; Smithson & Mollon, 2006). Symbolic representations transform the subsymbolic information into representations of contours, shapes, and surfaces.

In biological vision, our initial encoding has to do with activations of units sensitive to oriented contrast that are not organized together into contour tokens (Hubel & Wiesel, 1962). Symbolic recoding is needed for these activations to be perceived as a unified boundary defining a shape outline and then to recode regions into constant curvature primitives. In the current model, we focus on the second transition, going from a set of curvature points to a much smaller set of constant curvature regions. The set of curvature points is already symbolic in the sense that it defines of a property of a contour token but is much less abstract than the ultimate constant curvature representation. In the following chapter, we develop a biologically plausible implementation of the constant curvature model that goes from truly subsymbolic activations to a symbolic description of shape.

The output of the constant curvature model is an object-centric representation that is invariant to rigid planar transformations. It is also substantially compressed with regard to the number of bits of information in the representation. Typical shape contours often consist of thousands of points along the contour boundary. The abstract shape representation consists of much fewer (generally less than 30) constituent primitives.

Compression in the constant curvature model depends on its two free parameters: the size of the integration window with which the contour is segmented, and the minimum curvature difference needed for two CC primitives to be represented separately. Together, these parameters determine the model's balance between representing the original contour with high fidelity and encoding a representation that is simple and invariant to contour changes that are undetectable in human perception. In Experiments 1 and 2, we designed psychophysical experiments to estimate these two parameters. We hypothesized that the visual system represents shapes with the coarsest possible parameters at which curvature changes and curvature variability are not detected in human perception.

In Experiment 1, we tested subjects' sensitivity to differences in curvature. We created a contour fragment made up of two smoothly connected constant curvature segments and asked subjects to judge its complexity relative to a contour fragment made of just one constant curvature segment. We expected than any contour fragment that was encoded with primitives would be judged to be more complex than the single curvature contour, but that for stimuli with two curvatures that were sufficiently similar, subjects would encode the whole fragment with a single curvature value and perceive it to have equal complexity with the stimulus truly made up of a single curvature.

The results revealed a psychometric function wherein subjects performed at chance picking out the more complex contour when the curvature ratio between segments was small, then improved monotonically up to a curvature ratio of about 1.3:1, after which increasing the curvature difference did not significantly improve performance. We fit a logistic regression to the data and found the 75% threshold at a curvature ratio of 1.18:1. This is the ratio we used to

fix the minimum curvature difference between adjacent segments in the constant curvature model.

Participants in Experiment 1 were never given any instruction about how to judge contour complexity. Still, they had no confusion picking out the stimulus made up of two segments as more complex in any trial besides the ones in which the difference was extremely small. This subjective notion of contour complexity differs from previous theories in which complexity depends on the amount of curvature within a contour (Attneave, 1954; Feldman & Singh, 2005; Norman, Phillips & Ross, 2001). Under this view, whether the two-segment contour or the one-segment contour was more complex should depend only on whether the curvature in the one-segment contour was higher than the mean of the curvatures in the two-segment contours. In Feldman and Singh's (2005) derivation, a straight-line continuation of the contour in the direction tangent to the last point adds the least new information. Participants' subjective reports in Experiment 1 seem to suggest that a continuation in the last point's curvature adds the least amount of new information.

The notion that contour complexity depends on the amount of curvature in a contour has been supported experimentally by studies in which subjects were asked to pick out the 10 most important points which, if connected by straight lines, would give the most accurate representation of the shape. Subjects reliably picked the points of highest curvature (Attneave, 1954; Norman et al., 2001). However, the results of these experiments might be different if subjects were told to pick points to be connected by segments of constant curvature. In that situation, subjects would be best served by picking the 10 points of highest curvature *change* rather than highest absolute curvature. This is similar to what we observe in Experiment 1, where

subjects judge the stimuli with one curvature change as more complex than the stimuli with none.

In Experiment 2, we tested subjects' sensitivity to curvature variation within a contour region. In the constant curvature model, this is determined by the size of an integration window that convolves over a 2D contour and divides the boundary into regions of similar curvature. If the window is small, the model has low tolerance for curvature variation and divides the contour into many small pieces. As the window gets larger, it abstracts over more curvature variation, dividing the contour into fewer pieces. We estimated the integration window size that most closely corresponds to shape encoding in human perception by comparing subjects' ability to discriminate between a contour with no constant curvature and a constant curvature representation of the contour produced from various window sizes.

Subjects performed at near-chance levels for small window sizes, indicating that the constant curvature representation of the shape was at least as precise as subjects' internal representation. For larger window sizes, subjects showed monotonic improvement in the discrimination task. We analyzed these data with a continuous piecewise linear regression and found no evidence for a non-zero slope with a window size up to 1.29 degrees of contour length, but strong evidence for a positive slope for larger window sizes. We hypothesized that the visual system will use the largest window size at which the constant curvature representation is not distinguishable from the original contour, so we fixed it at this transition point.

Subjects' difficulty in distinguishing between contour fragments and constant curvature representations of the fragments gives some evidence that constant curvature primitives are used in human shape encoding. Other primitives like straight lines (e.g., Pavlidis, 1982; Gdalyahu, 1999) could also approximate the contour well enough to be indistinguishable from the original,

but likely with far more components. Constant curvature representations that cannot be discriminated from the original contour are still relatively economical, generally consisting of eight to 12 segments.

A possible problem with specifying the integration window in terms of degrees of visual angle is that model outputs will change with scale. Difficulty arises, however, in specifying the size of an integration window in a way that is both scale invariant and unaffected by fragmenting or partially occluding a contour. One solution could be to treat occluded contours and contour fragments as pieces of a whole contour, even if the whole contour is unseen. Rather than use an absolute integration window size, we specified the window size as a percentage of a closed shape contour. For open contours like the stimuli used in Experiment 2, we computed window size as a percentage of the contour fragment, then computed the percentage of the whole shape contained in the fragment. Since closed contours always have a sum turn angle of 360 degrees, we could estimate the percentage contained in a fragment by computing the sum turn angle of the open contour divided by 360 degrees. Using this correction, we found an object-centric integration window size from the data in Experiment 2, about 4% of a closed shape's contour.

An alternative solution for finding a scale-invariant window size might be to generate a description of object shape using biologically inspired local edge detectors. Classic work in neurophysiology has found that the visual system is specially tuned to straight, oriented luminance contrasts in various positions in visual field (Hubel & Wiesel, 1962). Rather than think of curvature in a mathematical sense, it can be considered as the turn angle between two contrast detectors connected end-to-end along a contour. Crucially, oriented edged detectors operate across different scales (e.g., Ringach, 2002; Sachs, Robson & Nachmias, 1971). Different shapes might be best captured by edge detectors of different scales. For example, a

shape with very high curvature likely needs small detectors to capture the rapid change along the contour, while a shape with low curvature might be captured almost equally well by larger detectors.

If we hypothesize that the visual system uses the largest scale detectors that adequately capture the contour's behavior, scale invariance naturally falls out of the constant curvature model: two shapes that differ only scale have the same number of detectors and the same turn angle between detectors, differing only in the size of the detectors (for more, see Kellman & Garrigan, 2007). In this construction, a constant curvature primitive is encoded not by its curvature and arclength, but by its turn angle and the number of detectors in the segment. Returning to the issue of integration window size, the window size might be specified neither by absolute distance, nor a percentage of the closed contour, but as a fixed number of detectors to be considered. In small shapes, the curvature is higher and the detectors will be smaller, while in large shapes the curvature will be lower and the detectors will be larger. Both shapes, however, will have the same number of detectors, and if the integration window is based on detectors considered, the segmentation will be the same.

The results of Experiments 1 and 2 experimentally fixed both free parameters in the constant curvature model. In Experiment 3, we looked for evidence that the model outputs with these parameters resembled human subjects' internal shape representations. As a test, we generated shape pairs and computed the number of CC segments in each member of the pair. We then sorted the pairs into five categories based on the difference in segments between the shapes in the pair. Importantly, shapes in all five categories were equated in terms of physical contour difference; the amount of contour overlap between pairs in the zero-segment-difference category was identical to that in the five-segment-difference category (and in all other categories). In

structural models of shape, there are two possible kinds of shape differences. Qualitative shape differences can be thought of as a change in the number of shape primitives composing the object, such as the addition or deletion of a constant curvature segment in our model, or of an axial branch in skeletal shape models. Metric shape differences are changes in the features of a shape primitive (see Briscoe (2008) for discussion). In our model, these might correspond to changes in the curvature or angular extent of one of the segments. The same amount of physical shape change can be achieved by different amounts of qualitative and metric shape changes. We expected that even though these changes give rise to the same amount of contour differences, qualitative changes (i.e., changes in the number of segments in a representation) will produce more perceptually different shape representations than metric changes.

We therefore predicted that, if the constant curvature model's outputs are similar to abstract shape representations, subjects should be better at detecting differences between shape pairs that had a greater segment number difference than shape pairs that differed by fewer segments. The results of Experiment 3 appear to validate the parameterized model, showing a clear linear trend in which shape pairs with smaller differences in CC segment number were more difficult to distinguish than shape pairs with larger segment number differences. The model is able to identify contour differences that will be perceptually salient, even when those differences do not correspond to larger changes to the physical contour.

The parameters we estimated in Experiment 1 and 2 appear to reflect human perceptual performance well for shape recognition. It is an open question whether the visual system's representational fidelity varies with perceptual task. Possibly, if people were encoding a shape representation for a task that required very high precision, some of the curvature variation that appears to be abstracted over in our experiments would be preserved, at least for as long as

subjects could hold the representation in working memory. On the other hand, as we have argued, it seems likely that even when the visual system encodes higher fidelity representations for specific perceptual task, a basic shape representation used for recognition will also be encoded with the parameters estimated in Experiments 1 and 2. Without this basic, task-independent representation, recognition would only be possible when shapes were viewed under similar task conditions.

Another open question is whether the complexity of the stimulus affects encoding precision. It seems possible that shapes with many parts and highly variable curvatures might be encoded more coarsely than simple shapes where perceptual resources are available to capture more subtle curvature variations. The results of Experiments 1-3, however, suggest that this might not be the case. In Experiments 1 and 2, we used simplified stimuli to estimate the model's parameters. If the visual system takes advantage of this simplicity to encode higher-fidelity representations, then their predictive power on more complex stimuli, like those used in Experiment 3 should be fairly weak. We cannot rule out the possibility that parameters fixed with stimuli of similar complexity to those used in Experiment 3 might be more predictive of perceptual performance, but it seems more likely that the precision of shape encoding does not depend on contour complexity.

Although the findings of our experiments lend evidence for constant curvature shape processing, we must note that this model does not exclude other theories of shape, such as skeletal or structural models. The constant curvature model aims to clarify how initial, subsymbolic boundary points are recoded abstractly. For some kinds of shapes, further processing might be needed to provide invariance under articulation, a task to which skeletal models are well-suited, or to allow recognition for objects in the same basic category (Rosch,

Mervis, Gray & Johnson, 1976), a task better suited to structural descriptions (e.g., Biederman 1987). We hypothesis that any further processing is done on the outputted constant curvature representation, not the subsymbolic contour description.

Conclusion

In this study, we have outlined a model of object shape representation that takes boundary points as inputs and produces an abstract representation made up of constant curvature primitives. The outputted representation is much more informationally compact than the activating stimulus, while still including perceptual features that appear to be essential in shape encoding and recognition. Representations from this model are also object-centric, invariant to planar transformations, and robust to fragmentation and partial occlusion. In Experiments 1 and 2, we fixed free parameters that govern how the visual system might balance constraints for simplicity and fidelity. We then validated the parameterized model in Experiment 3, by testing a behavioral prediction about how shape pairs matched for physical similarity might be perceptually more similar or different. Taken together with other evidence of the importance of constant curvature in visual processing, these findings provide strong evidence for the encoding of constant curvature representations of shape from object boundaries.

References

Altmann, C. F., Bülthoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology*, *13*(4), 342-349.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, *61*(3), 183.

Ayzenberg, V., & Lourenco, S. F. (2019). Skeletal descriptions of shape provide unique perceptual information for object recognition. *Scientific Reports*, *9*(1), 9359.

Baker, N., & Kellman, P. (2017). Psychophysical Investigations into Skeletal Shape Representations. *Journal of Vision*, *17*(10), 1379-1379.

Baker, N., & Kellman, P. J. (2018). Abstract shape representation in human visual perception. *Journal of Experimental Psychology: General*, *147*(9), 1295.

Barenholtz, E., Cohen, E. H., Feldman, J., & Singh, M. (2003). Detection of change in shape: An advantage for concavities. *Cognition*, *89*(1), 1-9.

Bertamini, M., Palumbo, L., & Redies, C. (2019). An advantage for smooth compared with angular contours in the speed of processing shape. *Journal of Experimental Psychology: Human Perception and Performance*.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, *94*(2), 115.

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, *20*(5), 585-593

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 121.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive psychology*, *20*(1), 38-64.

Blum, H. (1973). Biological shape and visual science (Part I). *Journal of theoretical Biology*, *38*(2), 205-287.

Briscoe, E. (2008). *Shape skeletons and shape similarity*.

Buffart, H., Leeuwenberg, E., & Restle, F. (1983). Analysis of ambiguity in visual pattern completion. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(6), 980.

Chow, C. C., Jin, D. Z., & Treves, A. (2002). Is the world full of circles?. *Journal of vision*, *2*(8), 4-4.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & psychophysics*, *27*(3), 183-228.

Cooper, L. A., Schacter, D. L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 43.

Elder, J. H., & Velisavljević, L. (2009). Cue dynamics underlying rapid detection of animals in natural scenes. *Journal of Vision*, *9*(7), 7-7.

Elder, J. H., Oleskiw, T. D., Yakubovich, A., & Peyré, G. (2013). On growth and formlets: Sparse multi-scale coding of planar shape. *Image and Vision Computing*, *31*(1), 1-13.

Farah, M. J., Rochlin, R., & Klein, K. L. (1994). Orientation invariance and geometric primitives in shape recognition. *Cognitive Science*, *18*(2), 325-344.

Fechner, G. T. (2012). Elemente der psychophysik. Рипол Классик.

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological review*, *112*(1), 243.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014-18019.

Garrigan, P. (2006). *Representation of contour shape*.

Garrigan, P., & Kellman, P. J. (2011). The role of constant curvature in 2-D contour shape representations. *Perception*, *40*(11), 1290-1308.

Garrigan, P., Baker N., & Kellman, P.J..(2020). Constant Curvature Segments as Building Blocks for 2D Shape Representation. Under review in *Journal of Experimental Psychology: General*.

Gdalyahu, Y., & Weinshall, D. (1999). Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12), 1312-1328.

Hatfield, G., & Epstein, W. (1985). The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin*, *97*(2), 155.

Hochberg, J., & McAlister, E. (1953). A quantitative approach, to figural" goodness". *Journal of Experimental Psychology*, *46*(5), 361.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, *18*(1-3), 65-96.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106-154.

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive development*, *9*(1), 45-75.

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of neurophysiology*, *73*(1), 218-226.

Kanizsa, G. (1976). Subjective contours. *Scientific American*, *234*(4), 48-53.

Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision*, *1*(4), 321-331.

Kellman, P. J., & Garrigan, P. (2007). Segmentation, grouping, and shape: Some Hochbergian questions.

Kellman, P. J., Garrigan, P., & Erlikhman, G. (2013). Challenges in understanding visual shape perception and representation: Bridging subsymbolic and symbolic coding. In *Shape Perception in Human and Computer Vision* (pp. 249-274). Springer, London.

Kimia, B. B., Tannenbaum, A. R., & Zucker, S. W. (1995). Shapes, shocks, and deformations I: the components of two-dimensional shape and the reaction-diffusion space. *International journal of computer vision*, *15*(3), 189-224

Koffka, K. (2013). *Principles of Gestalt psychology*. Routledge.

Kovacs, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences*, *90*(16), 7495-7497.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive development*, *3*(3), 299-321.

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics*, *80*(5), 1278-1289.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279.

Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex*, *4*(5), 523-531.

Norman, J. F., Phillips, F., & Ross, H. E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception*, *30*(11), 1285-1294.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of neurophysiology*, *86*(5), 2505-2519.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature neuroscience*, *5*(12), 1332.

Pavlidis, T. (2012). *Algorithms for graphics and image processing*. Springer Science & Business Media.

Pizlo, Z., & Stevenson, A. K. (1999). Shape constancy from novel views. *Perception & Psychophysics*, *61*(7), 1299-1307.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(3), 422.

Prins, N. (2009). FAA Kingdom,"Palamedes: Matlab routines for analyzing psychophysical data".

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of neurophysiology*, *88*(1), 455-463.

Rock, I., & Epstein, W. (1977). Stability and constancy in visual perception: Mechanisms and processes.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382-439.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America*, *61*(9), 1176–1186.

Salkauskas, K., & Lancaster, P. (1990). Curve and surface fitting.

Schmidt, F., Spröte, P., & Fleming, R. W. (2016). Perception of shape and space across rigid transformations. *Vision research*, *126*, 318-329.

Sigman, M., Cecchi, G. A., Gilbert, C. D., & Magnasco, M. O. (2001). On a common circle: natural scenes and Gestalt rules. *Proceedings of the National Academy of Sciences*, *98*(4), 1935-1940.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2008). Are there multiple visual short-term memory stores?. *PLOS one*, *3*(2), e1699.

Smithson & Mollon, 2006

Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and applied*, *74*(11), 1.

von Helmholtz, H. (1962). *Helmholtz's treatise on physiological optics* (Vol. 3). Dover Publications.

Weber, E. H. (2018). EH Weber on the Tactile Senses. Psychology Press.

Wu, H., Frund, I. & Elder, J. (2016). What is Perceptual Curvature?. *Journal of Vision*, *16*(12), 812-812.

Wuescher, D. M., & Boyer, K. L. (1991). Robust contour decomposition using a constant curvature criterion. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (1), 41-51.

Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology*, *49*(2), 155-190.

Chapter 2: A Biologically Plausible Model for Constant Curvature Shape Representation

Nicholas Baker[1], Patrick Garrigan[2] & Philip J. Kellman[1]

1. Department of Psychology, University of California, Los Angeles
2. Department of Psychology, Saint Joseph's University

Abstract

A model of constant curvature representation is proposed that bridges the gap between subsymbolic and symbolic descriptions of a shape stimulus. The model begins with outputs from local oriented contrast detectors at twelve orientation channels. A new kind of detector called arclets is then proposed which are tuned to a specific turn angle between oriented detectors. The true curvature of points along a shape's boundary is estimated by the population code of various arclets at a given spatial position. The model then uses curvature information to partition the shape's contour into regions of similar curvature. This process is both local and invariant to scale. Finally, the shape is recoded symbolically as a set of constant curvature segments that each have a specified curvature and angular extent.

In the previous chapter, we developed a model for how the visual system represents a shape contour with relatively few smoothly joined constant curvature segments. The proposed model was formulated at the computational level of description (Marr, 1982), taking a set of contour points as input and outputting an organization of similar curvatures into distinct segments. In this chapter, we will aim to go a step further and propose a more biologically inspired model for constant curvature shape representation that works at the algorithmic level of description.

The proposed model, which we call the detector model, aims to more fully bridge the theoretical gap between subsymbolic and symbolic visual representations. As discussed in the previous chapter, subsymbolic representations refer to neural activations in early visual areas that are primarily sensitive to energy contrasts. Though they are essential to vision, the goal of perception is not to detect contrast energies, but to infer properties of objects and events in the world (Gibson, 1966, 1979; Marr, 1982). Symbolic visual representations, including descriptions of contours, surfaces, and shape, encode properties that often do not have inherent energy, but about which inferences can be made from patterns of energy contrast.

The constant curvature model developed in the previous chapter makes some progress in connecting subsymbolic and symbolic codes. In the current work, we more fully develop the subsymbolic side of the model, beginning from outputs of oriented edge detectors known to operate at various scales in V1 (Hubel & Wiesel, 1962; Hubel & Wiesel, 1968) and going to a wholly symbolic description of shape through progressive layers of abstraction. In doing so, we aim to connect work being done in research on early vision with higher level notions of perception that often do not interact with low level research.

In developing this model, we were motivated by four central goals. First, we wanted it to be biologically plausible. We recognize that orientation coding in early visual areas is extremely complex (Celebrini, Thorpe, Trotter & Imbert, 1993; Pei, Vidyasagar, Volgushev & Creutzfeldt, 1994; Ringach, Hawken & Shapley, 1997), and we do not fully grapple with all of its nuances, but we aimed to simplify these complexities only in ways that are consistent with prevailing opinions about how the visual system operates.

Second, we wanted the detector model to operate locally. By that we mean that the segmentation of part of a contour into constant curvature pieces should depend only on the features of that local region of the contour. In the computational model from the previous chapter, we used an integration window to compare curvatures in a local region, but the size of the integration window was a percentage of the total length of the contour. Consequently, if two shapes had the same local contour region, but one had a separate part that extended further than the other, the size of the integration window would differ for the two shapes, and the same local region would be segmented differently. Likewise, shapes in a real visual scene are often partly occluded, making it difficult to know what the contour's true length is. If the visual system inaccurately estimated the length of a partly occluded shape, it would segment two identical shapes differently if one was occluded and the other was not. These struck us as unlikely predictions, so we aimed to amend the model so that only local curvatures affected a contour region's segmentation.

Third, we wanted the detector model to be scale invariant. It is well established that two shapes that differ in scale are perceived as the same, for many tasks or purposes, in both perception (Biederman & Cooper, 1992; Cooper, Schacter, Ballesteros & Moore, 1992; Schmidtmann, Jennings & Kingdom, 2015; Baker & Kellman, 2018) and neurophysiology of

high level regions in visual cortex (Lueschow, Miller & Desimone, 1994; Ito, Tamura, Fujita & Tanaka, 1995). When the scale of an object changes, all curvature values along its contours also change such that if a shape's size is doubled, the variation in curvature values will be halved. In computer vision, one way to build in scale invariance is to normalize the representation by the contour's total length (Belongie, Malik & Puzicha, 2002). We did something similar in the computational version of our constant curvature model by fixing the integration window size as a function of the contour length so that the same percentage of the contour was considered for two matched shapes that differed in scale. Of course, fixing the integration window size in this way is the very thing that violates the locality constraint described above. We aimed to satisfy both constraints in the detector model.

Finally, we wanted the detector model to give outputs that were similar to those generated from the computational model. Experiment 3 in the previous chapter found that the outputs of the parameterized computational model were predictive of perceptual performance on a shape recognition task, even when the physical similarity of shape pairs was equated across conditions. We would not want to replace this model with a different one that has less predictive capability, even if it satisfied certain other theoretical constraints.

In the proposed model, we begin with a simple image of a white figure on a black background. We find activations of odd-symmetric gabors convolved over all positions in the image at 12 different orientations. We then propose a higher-order detector that is connected to spatially adjacent orientation detectors and is tuned to a specific turn angle. We show how the true curvature of a point along the contour can be estimated by the pattern of activations from detectors tuned to different turn angles at the same position in visual field. Next, we use the estimated curvatures at the finest scale to partition the shape into regions of similar curvature and

recode them into regions with constant curvature. Constant curvature segments are then assigned an appropriate scale based on their curvature, such that regions of low curvature are represented by larger scale detectors and regions of high curvature are represented by smaller scale detectors. Finally, we compute a symbolic code for the shape where each segment is described by a small number of abstract parameters.

Model

Input

The input to the model is a 512 x 512 pixel image with a black background and a filled in shape with uniform illumination. In most simulations, we used a white shape, but tests were also conducted on shapes with various luminance levels in grayscale. In all but the lowest contrast cases, outputs were the same regardless of luminance. An example input image is shown in Figure 1.

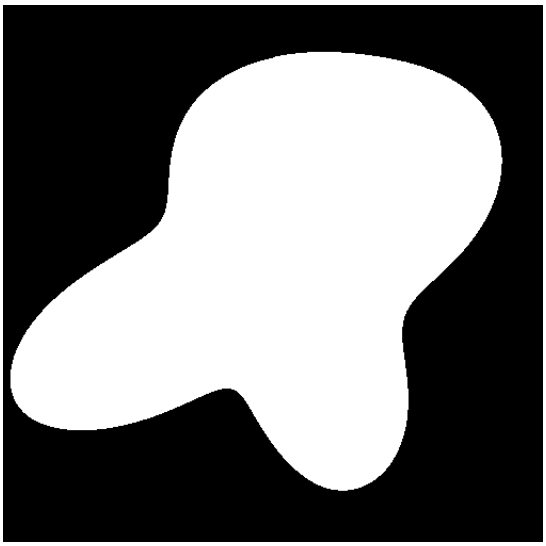

Figure 1. Sample input stimulus to the detector model.

Orientation Filters

The first stage of processing involved finding activation levels of oriented detectors for all positions along the contour. In vision research, Gabor filters have been used to model

activations of simple and complex cells in striate cortex (Kulikowski, Marcelja & Bishop, 1982; Daugman, 1985). We used a version of Gabors called S-Gabors developed by Heitger and colleagues that have the useful property of having zero activation in a field of uniform luminance (Heitger, Rosenthaler, Von Der Heydt, Peterhans & Kubler, 1992; Ludtke, Wilson & Hancock, 2000; Hickinbotham, Hancock & Austin, 1997). Because we were dealing with filled shapes, we took activations only from odd-symmetric Gabors that had an excitatory response to luminance on one side of the line of symmetry and an inhibitory response to luminance on the other. Following Heitger et al, 1992, we define the odd-symmetric S-Gabor as:

$$G_{odd}(x) = e^{-\frac{x^2}{2\sigma^2}} \cdot \sin\left[2\pi v_0 x \xi\,(\mathrm{x})\right] \qquad (1)$$

where σ determines the width of the Gaussian envelope and v_0 is the frequency at the origin. The difference from the simple odd-symmetric Gabor is given by the additional function $\xi\,(\mathrm{x})$, a frequency sweep defined as:

$$\xi\,(\mathrm{x}) = k \cdot e^{-\lambda \cdot \left(\frac{x}{\sigma}\right)^2} + (1-k) \qquad (2)$$

where *k* is the relative amplitude of the frequency sweep (from Heitger et al., we used *k* = 0.5.) and λ is chosen so that

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \cdot \cos[2\pi v_0 x \xi\,(\mathrm{x})] = 0 \qquad (3)$$

$\xi\,(\mathrm{x})$ changes the frequency selectivity slightly as distance from the receptive field center increases.The justification for this modification of the original Gabor function is that it gives zero response for a uniform field in quadrature pairs for all phases of filters, which may be a desirable characteristic both theoretically and empirically (see discussion in Heitger et al, 1992).

Since the S-Gabor is operating on a 2D image, we rewrite (1) and (2) so they define a receptive field in 2D space:

$$G_{odd}(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \cdot \sin\left(2\pi v_0(x \cdot cos\theta - y \cdot \sin\theta * \xi(x,y)\right) \quad (4)$$

$$\xi\,(\mathrm{x},\mathrm{y}) = k \cdot e^{-\lambda\cdot(\sigma\cdot(x\cdot\cos\theta - y\cdot sin\theta))^2} + (1-k) \qquad (5)$$

A sample odd-symmetric detector is shown in position space in Figure 2. The S-Gabor defined in (4) is circular. Following Heitger et al., we elongated the filters to more closely reflect the characteristic shape of detectors in striate cortex (Ringach, 2002). In human perception, the finest scale of the detector depends on its retinal eccentricity. Though we were primarily interested in modeling foveal detectors, size estimations for them are extremely difficult, as even tiny eye movements in anesthetized animals make recordings difficult (Ringach, personal communication). Detectors near the fovea (5 degrees of eccentricity) have a peak spatial frequency sensitivity of about 10 cycles per degree (Hawken & Parker, 1987). For our purposes, since the number of pixels in a degree of visual angle depends on screen size and viewing distance, we chose a slightly larger detector ($\sigma = 1.5$) as our finest scale to better smooth over pixilation effects in how the shape was digitally rendered. If the box enclosing the display in Figure 1 had physical dimensions of 13.54 cm by 13.54 cm on a 58.4 cm screen, and was viewed from 160 cm away, these filters would have a spatial frequency sensitivity of about 8.8 cpd. For such a stimulus rendered on a display screen at 512 by 512 pixels these filters would be approx. 38 pixels (length) by 24 pixels (width), corresponding at this viewing distance to 13min of visual angle in width.

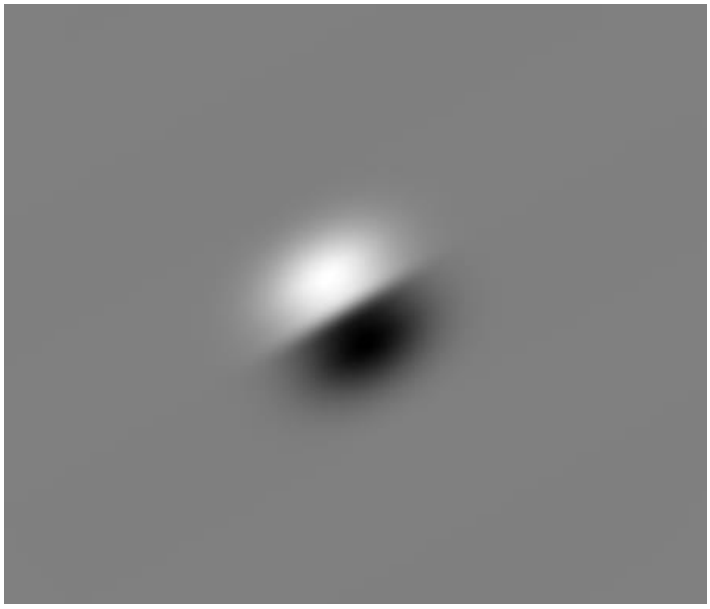

Figure 2. Sample odd-symmetric S-Gabor used in the detector model. The scale of the detector is increased for purposes of presentation.

Following Heitger et al. (1992), we convolved the S-Gabor in 12 orientation channels from zero degrees to 360 degrees in 30-degree increments. To get the true orientation at the level of sensitivity found in human perception (Clifford, Wyatt, Arnold, Smith & Wenderoth, 2001), activations of these broadly tuned neurons had to be combined into a population code (Georgopoulos, 1986; Adelson, 1991; Ludtke et al., 2000). We estimated the true orientation at each position by taking the orientation channels with the top three activation scores and computing

$$\theta_E = (\theta_1 * A_1 + \theta_2 * A_2 + \theta_3 * A_3)/(A_1 + A_2 + A_3) \qquad (6)$$

where Θ_E is the estimated orientation. $\Theta_{1\text{-}3}$ are the three orientation channels with the highest activation, and $A_{1\text{-}3}$ are the activations of each of the top three orientation channel. A comparison between the analytically computed orientation and the orientation estimated by population coding from the 12 detectors for a part of the shape is shown in Figure 3.

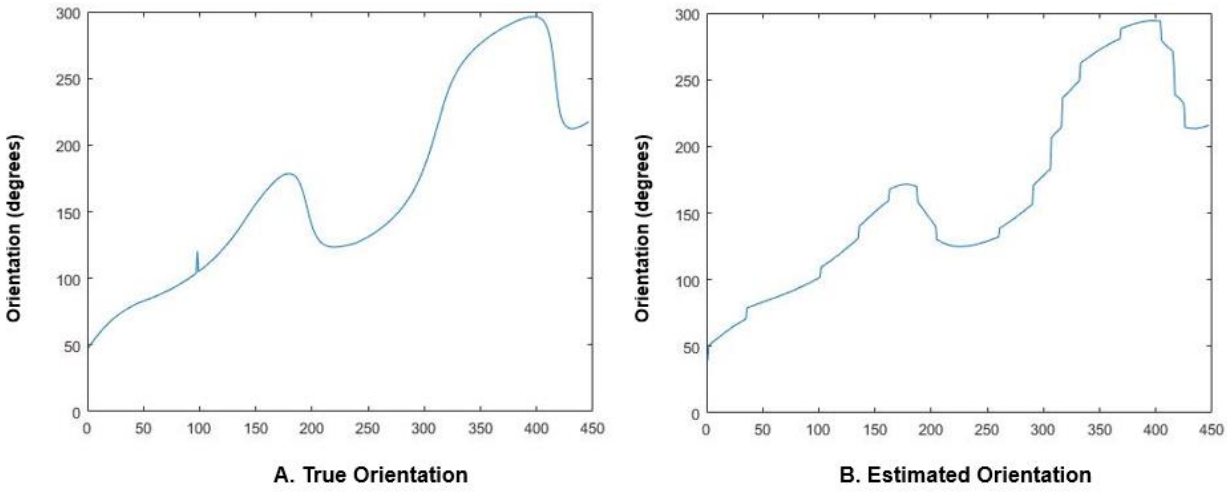

Figure 3. Comparison between analytically computed orientation (A) and orientation estimated by population coding from S-Gabors (B). The analytically computed orientation was computed by connecting straight lines between points along the contour and computing the orientation of the lines relative to the reference frame.

Curvature Estimation

The goal of our model is to segment a contour into regions of similar curvature. In the computational version of our model, signed curvature (c) was computed from the contour points analytically with the equation

$$c = \frac{x'y'' - y'x''}{(x'^2 y'^2)^{3/2}} \qquad (7)$$

We shifted the detector model from this symbolic computation to something that can be computed from the outputs of oriented detectors. A more biologically plausible analogue for curvature is the turn angle between adjacent oriented detectors along a contour (Kellman & Garrigan, 2007). In the limit, as the scale of adjacent detectors becomes infinitesimally small, this curvature estimation will approach the analytically computed true curvature of the contour.

One way we considered estimating curvature by change in orientation was to take the estimates of true orientation we computed for points along the contour (Figure 3) and compute the difference in orientation for all adjacent points. Doing so would give accurate estimations of

curvature, but it also constitutes a change from a subsymbolic code to a symbolic code, as there must be a single unit that is tuned to every possible orientation for each point along the contour. Research on population codes suggests that individual neurons tuned to all possible continuous values are unlikely (Ma, 2010).

An alternative way of computing curvature would be to create *arclets* that are connected to pairs of oriented detectors and estimate a continuous curvature value from the population code of their activations in a given position in visual field. In our model, there are 12 oriented channels for each spatial position, so there are 144 arclets for each pair of adjacent edge positions corresponding to all possible pairings of orientation channels. Each arclet, then, is tuned to a single turn angle in the same way that S-Gabors are tuned to a single orientation, where turn angle is simply the difference in the preferred orientation of the two detectors to which the arclet is connected.

We computed "curvature columns" consisting of the 144 arclets spaced around the shape's contour. Each arclet was comprised of two oriented detectors in coaxial spatial positions, one whose center was clockwise along the contour and one whose center was counterclockwise along the contour. The spacing between arclets was such that the Gaussian envelope of one arclet's clockwise detector overlapped with (and only with) the Gaussian envelope of the next clockwise arclet's counterclockwise detector (see Figure 4).

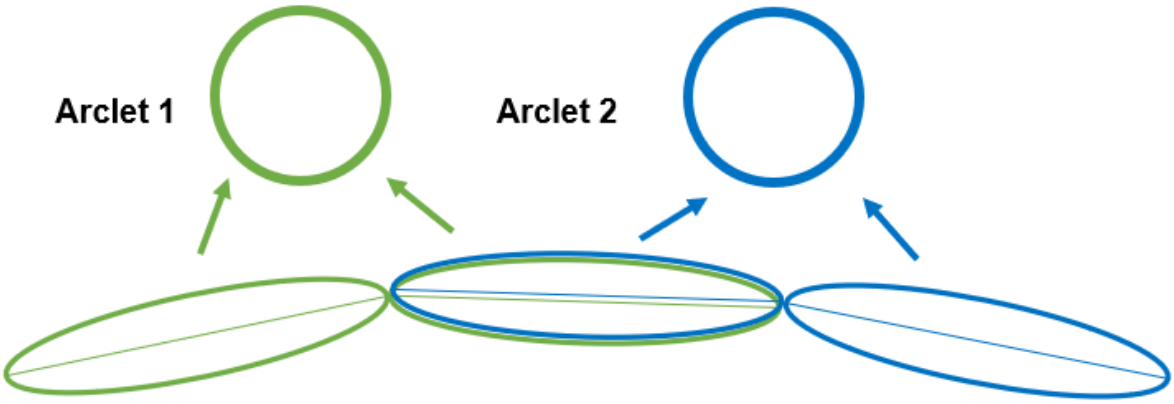

Figure 4. Arclets and oriented detectors. Here, we show an example of two arclets in adjacent spatial positions along the contour. Not shown is that for each arclet position, there should be not one but 144 possible arclets in a curvature column.

For each arclet position along the contour, we had the activation of the 144 arclets in our curvature column. Activation for a particular arclet in the column was determined by the activation of the oriented detectors to which it was connected. For simplicity, we calculated the product of the activations for each detector. We eliminated arclets that had high activation from the product of two negative activations of the oriented with the simple rule that if either detector had negative activation, the arclet activation was always set to zero. We also tried adding the arclet activations together which worked about equally well, but we ultimately preferred to use the product so that arclets with pairs of detectors that both had strong activation would be more responsive than arclets whose activation was driven by only one of the detectors.

Once we had computed the activation for each arclet in the curvature column, we estimated the true turn angle of the contour using a population code from the activation of the 144 arclets. Many sophisticated algorithms have been developed to decode a continuous response from activations of populations of cells (e.g., Chen, Geisler & Seidemann, 2006; Berens, Ecker, Cotton, Ma, Bethge & Tolias, 2012; Graf, Kohn, Jazayeri, & Movshon 2011).

Here, because we are dealing with idealized S-Gabors and because the stimulus is greatly simplified to a single white figure on a black background, we simply used a weighted combination to estimate the turn angle. We found the top seven arclet activations at a given position and computed

$$\phi_E = \frac{\sum_1^7 \phi_{i^*} A_i}{\sum_1^7 A_i} \qquad (8)$$

where ϕ_E is the estimated turn angle, ϕ_i is the turn angle of the seven arclets with the highest activation, and A_i is the activation of the seven arclets with the highest activation.

A comparison of the true curvature profile for the shape in Figure 1 and the curvature profile estimated from population coding of arclets is shown in Figure 5. Although the global trends are very similar in both plots, there is considerable noise in the curvature estimation. One source of this noise is from aliasing along the shape's border. Figure 6 shows a part of the shape zoomed in. To get precise estimates of curvature, we use relatively small detectors, which results in more noise being introduced by these pixilation effects. Another possibility is that our simple calculation of arclet activation as the product of two detectors introduces undue influence from arclets that should have relatively low activation. We plan to consider more sophisticated ways of combining the detector activations in future simulations.

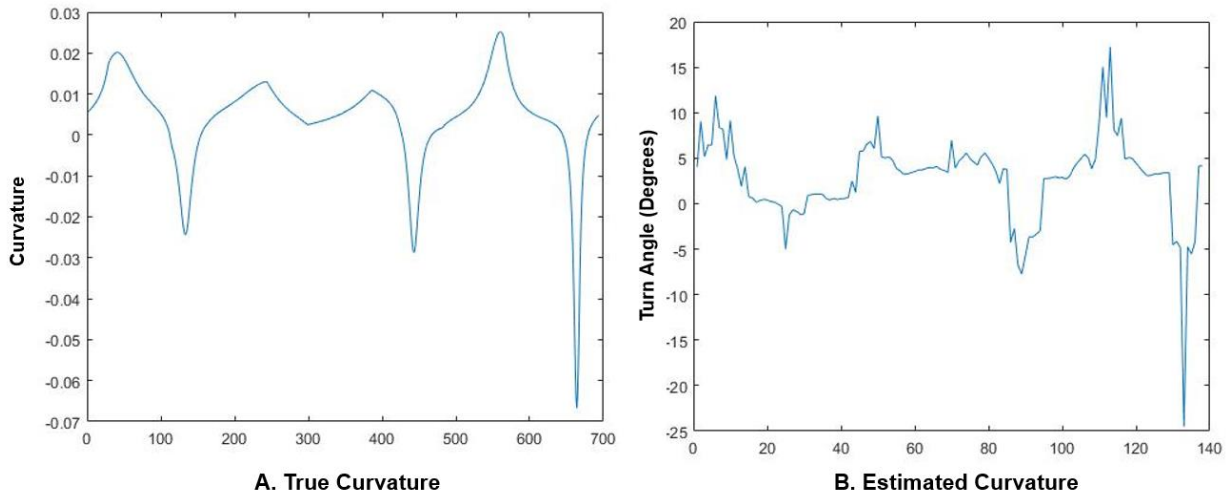

Figure 5. Comparison of the shape in Figure 1's true curvature (A) and estimated curvature from population coding (B).

Figure 6. Figure 1 zoomed in to show aliasing.

Constant curvature segmentation

After estimating curvature profiles from population codes of arclet activations, we developed a local and scale invariant algorithm for segmenting nearby curvatures into segments that can be approximately described by a unique curvature value. In the algorithm, contour

regions of similar curvature are successively organized together until no pair of adjacent regions have a curvature difference below the model's threshold.

In the simulations reported in this section, we do not use the estimated curvature from population coding, but the true curvature calculated analytically. Ultimately, the goal of this model is to link up subsymbolic curvature descriptions with constant curvature segmentation, but we felt it would be easier to work with less noisy curvature estimates while developing the segmentation algorithm. The results of curvature estimation from population coding look promising and we are confident that local noise in the estimated curvature profile can be reduced, but until then we report findings using the true curvature profile.

The algorithm begins with the assumption that differences in curvature polarity always signal a segment boundary. A coarse initial segmentation of the shape is given by the set of contour points between zero crossings in the shape's curvature profile. We then iteratively compared curvature for adjacent regions within a segment of matched polarity, combining regions with similar curvature together and drawing segment boundaries between regions with different curvature.

This second, more precise, segmentation organized adjacent contour regions together if they satisfied one of two basic rules. The primary rule was that if the two regions' difference in curvature was below some threshold, *T*, they were organized into a single segment. Based on findings from Experiment 1 in the previous chapter, the visual system appears to be sensitive to ratios of curvature differences, not absolute differences. We therefore fixed *T* as the proportional difference in curvature between the two segments. Formally, if $\left|\frac{(k_i - k_i + 1)}{k_i}\right| < T$, where k_i is the curvature at the i^{th} point along the contour and k_i+1 is the curvature one point further along the contour, then the two regions are organized into a single segment.

We included a secondary rule for organizing points with very low curvature together, since a large proportional difference between them could still reflect a very small absolute difference. To preserve scale invariance, we analyzed the curvature profile for the fifth percentile of curvatures. Research in perception has shown that the visual system can quickly compute statistical features of a display (e.g., Sun, Chubb, Wright & Sperling, 2018), so we do not consider it implausible that the visual system could quickly assess the bottom percentile of curvatures. Any adjacent segments whose curvatures were both below the fifth percentile were organized together, even if their proportional difference exceeded the threshold from the primary segmentation rule.

After two curvature regions were organized together, they were assigned a new curvature which was the weighted mean of the two segments' curvatures. Curvatures were weighted by the number of detectors at the finest scale they had in their respective segments. The algorithm begins by combining individual points at the smallest scale, but as these points are organized into larger segments, it continues to merge together adjacent segments until the ratio of the curvatures of all adjacent segments is larger than *T*.

Parameterizing and evaluating the segmentation algorithm

The only free parameters in the segmentation algorithm are the proportional difference in curvature below which contour regions are represented together, and the curvature percentile under which contour regions are represented together regardless of their proportional curvature difference. The curvature percentile is useful for preventing the addition of many spurious segment boundaries in areas of low curvature along the contour, but its overall effect on the shape's segmentation is small. We selected 5% arbitrarily, but simulations in which it was lowered to 1% or increased to 10% had little influence on the model's outputs in most cases.

The more important parameter in the detector model is the threshold for proportional curvature differences, which determines the model's degree of tolerance for curvature variety within a single segment. In practice, *T* serves the role that the integration window size and difference threshold serve together in the computational model. Consequently, the parameters experimentally estimated in the previous chapter are not directly applicable to the detector model. However, since we found behavioral evidence that the model outputs produced by the computational model with experimentally estimated parameters were predictive of human perceptual performance, we wanted the detector model's outputs to have similar levels of efficiency and precision.

We began by using outputs of the parameterized computational model to look for the threshold value in the detector model that most closely matched it in efficiency (i.e., in the number of segments in the outputted representation). We considered four different proportional difference thresholds, 50%, 55%, 60%, and 65%. Smaller thresholds correspond to more precise shape representations. As thresholds get larger, the model has greater tolerance for curvature variety and representations become simpler with fewer segments. We generated 20 shapes for each condition and applied both models to the same 20 shapes. In Figure 7, we plot the average difference in number of segments for the parameterized computational model as a function of the proportional difference threshold in the detector model. There was a strong linear relationship between the proportional difference threshold applied to the detector model and the number of segments by which the detector model differed from outputs of the parameterized computational model. We used this linear fit to estimate that the two models would, on average, be described by the same number of segments when the proportional difference threshold was 55.7%.

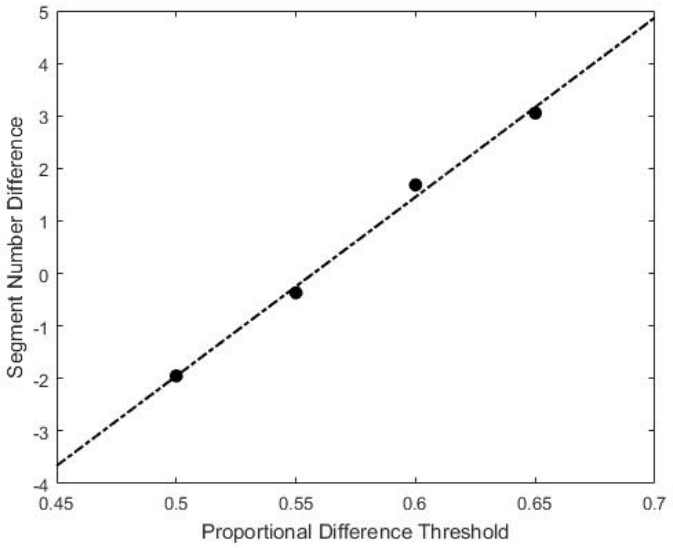

Figure 7. Relationship between proportional difference threshold and the average difference between constant curvature segments in the computational model and constant curvature segments in the detector model. The difference was averaged over 20 shapes and computed by subtracting the number of segments in the detector model from the number of segments in the parameterized computational model. The dashed line shows a linear fit for the data.

Once we had fixed the difference threshold in the detector model, we could compare the precision of the detector model with the precision of the computational model. Both models had, on average, the same number of segments in their representation, but it remained to be seen if one algorithm was more successful at segmenting the shape into regions of similar curvature. To assess this, we identified the segment boundaries placed on the same shape for the computational and detector model and measured the variance in the true curvature of the shape in each segment. We generated 30 shapes and computed the average variance in each segment for both models. The results are plotted in Figure 8. On average, the ratio of mean variance of the computational model to the detector model was 1.004:1.

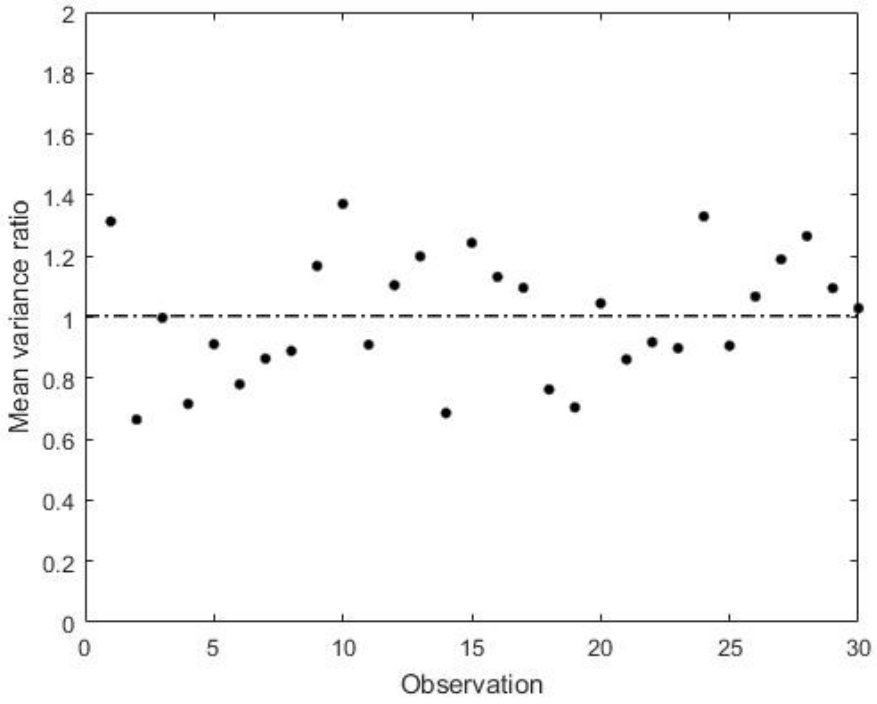

Figure 8. Ratio of the average curvature variance in outputs from the computational model to outputs in the detector model. Each point is the ratio for one of the 30 shapes in our test, and the dashed line show the mean ratio across all 30 shapes.

Figure 9 shows a shape stimulus and constant curvature representations of the shape produced by the computational model and the detector model side-by side. The results of our simulation suggest that constant curvature shape representations produced by the detector model are very similar to outputs from the computational model developed in the previous chapter. We fixed the threshold parameter in the detector model using parameters we had fixed experimentally in the computational model to equate the average number of segments produced for a shape in both models. After doing so, we found that the average precision of both models was equated as well, suggesting a general equivalence between the two systems. This satisfies one of the initial constraints we had set for the detector model.

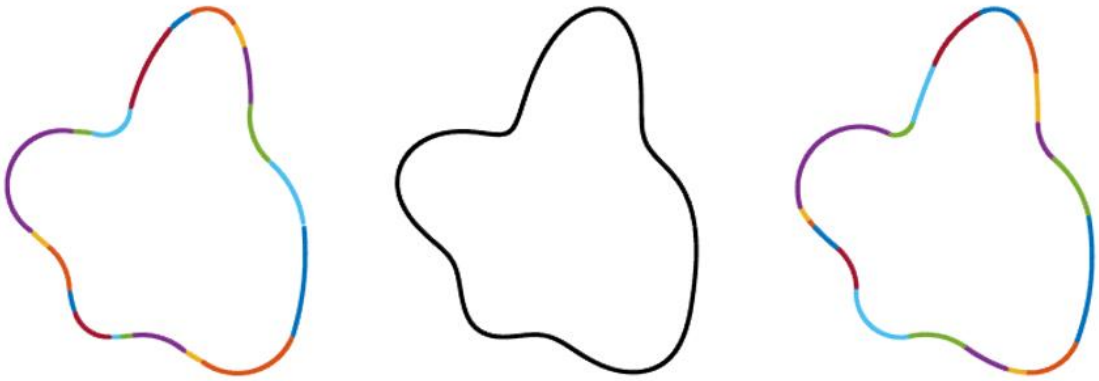

Figure 9. Constant curvature shape representation from the parameterized computational and detector models. Left: Output from the computational model. Middle: True shape. Right: Output from the detector model.

Another model constraint was that the segmentation should depend only on local curvature features. The detector model satisfies this constraint much more successfully than did the computational version of the model. Whereas the computational version included an integration window whose size depended on the overall length of the shape's boundary, the detector model does away with the integration window and iteratively builds larger constant curvature segments based on the curvature ratio of neighboring regions along the contour. Consequently, stretching, compressing, or occluding non-neighboring regions has no effect on the model's local curvature segmentation.

The model is also scale invariant in the sense that its constant curvature segmentation does not change depending on the size of the shape. If the scale of a shape was doubled, all its curvatures would be halved, but the ratio of curvatures would remain the same. Likewise, we fixed the minimum curvature below which regions are organized together regardless of curvature ratio as a percentile of the curvature distribution. If all curvatures change by the same amount, the contour points in the fifth percentile will not differ.

Multiscale Representation

In our evaluation of the detector model's segmentation algorithm, we claimed that the detector model was scale invariant. This is true in one sense, as the detector model will partition a shape into the same set of segments regardless of its scale, but some abstraction is still needed for the visual system to find equivalence between representations of shapes at different scales. For example, consider the two shapes shown in Figure 10. They are segmented into the same contour regions, but if we analyze the properties of corresponding constant curvature segments in the two shapes, they will have very different properties. Since segmentation is done with arclets at the finest scale, the larger shape will have twice as many curvature detectors in its segment. Likewise, the curvature assigned to a segment in the smaller shape will be twice as high as in the larger one.

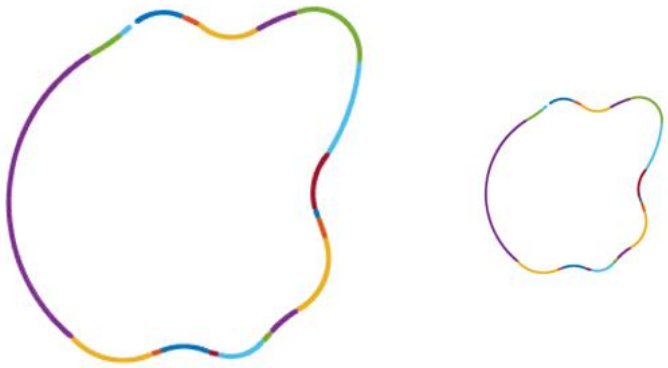

Figure 10. Two shapes that differ by a scale factor of 2.

It has been proposed that the visual system gets scale invariant descriptions of shape by representing arclets at different scales (Kellman & Garrigan, 2007; Baker, Kellman & Garrigan, 2020). When estimating the curvature profile of a contour through population coding, we used the smallest scale of arclets because those activations gave us the closest estimate of a contour's true curvature. However, it seems plausible that for any scale of oriented detectors, there are arclets connected to adjacent pairs. In the visual system's ultimate representation of a shape, a

constant curvature segment might not be described by the number of arclets at the smallest scale and the turn angle between them, but by relations of larger-scale detectors. The scale of detectors in the visual system's ultimate description likely depends on the absolute size and curvature of the shape stimulus.

Following work from neurophysiology, curvature in our model is approximated by difference in orientation between straight detectors (Hubel & Wiesel, 1962). The best approximation of curvature with straight detectors will always be the finest scale; the approximation converges toward the curved segment as the length of the straight detectors approaches zero. However, in many cases, the approximation from larger detectors will be quite close to the true stimulus, even if it is never closer than the approximation from smaller detectors. The goodness of the description of larger scale detectors to a contour depends on the curvature of the contour in that region. For example, consider the two circle contours in Figure 11. The green circle is twice as large as the blue circle. When detectors of the same size are placed along each circle's edge, we find a reasonable fit along the larger circle, but a poor fit along the smaller circle. For a zero-curvature segment, detectors of any length less than or equal to the total segment length would give as good a description as smaller-scale detectors.

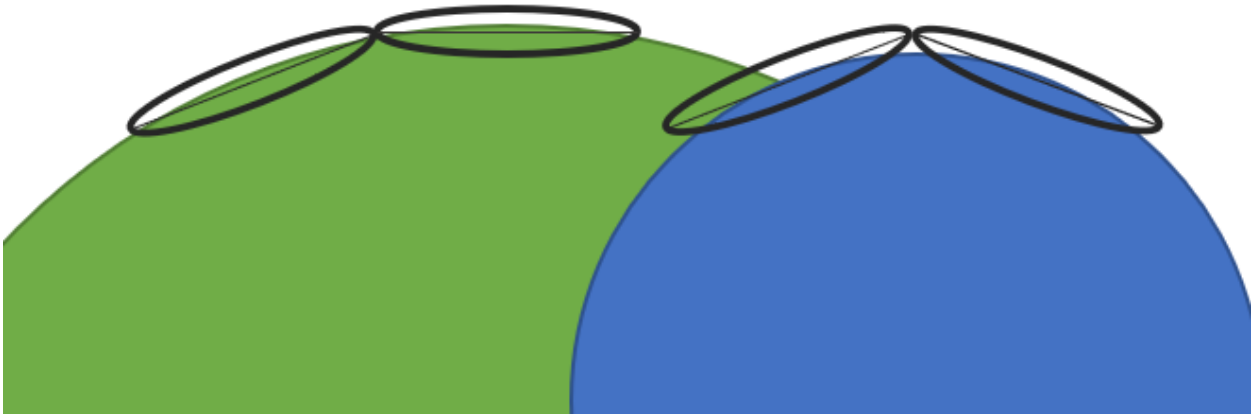

Figure 11. Oriented detectors placed along a circle of size S (blue) and 2S (green).

We propose that the visual system describes a constant curvature segment with the largest "good-enough" scale, by which we mean the largest-scale of detector that fits the stimulus with an error that falls below some error metric. One way to formalize this error is the area of the figure that falls in the inhibitory region of the detector plus the area of the ground that is in the excitatory region of the detector, normalized by the size of the detector. What the visual system's actual tolerance is for error when choosing a scale of the detector is an empirical question, but psychophysical research on the perception of curvature in straight lined polygons could shed some light on the problem.

The visual system might describe a region of the contour with the largest scale detectors that fit the actual curvature at or above some threshold tolerance. As discussed, this will result in large-scale detectors when curvature is low and small-scale detectors when curvature is high. When a shape is scaled up or down, the curvature of all segments increases or decreases by the same proportional amount. For example, consider the diagram in Figure 12. In the figure, two ellipses are shown that have identical form but a size ratio of 3:1. As discussed, the segmentation algorithm will partition both shapes into the same constant curvature primitives. Each primitive, then, will be described by a turn angle (Θ), a scale (k), and an extent (n). The scale variable refers to the size of the arclets that describe the contour. As Figure 12 shows, the largest-good-enough scale for an arclet in the ellipse of size 3k is three times larger than for the ellipse of size k. Crucially, though, the extent (i.e., the number of detectors in the segment) and the turn angle between detectors will remain the same across shapes of different scale. The relative scale of detectors between segments of different curvature could be represented as the ratio of the best-good-enough scale in the segment to the sum of all scales of segments throughout the shape. We

suggest that the visual system gets scale-invariant shape recognition by comparing segments' curvature (turn angle), angular extent (number of detectors), and relative size (normalized scale of detector), but not their absolute scale (see Baker, Garrigan & Kellman (2020) for empirical support of this hypothesis).

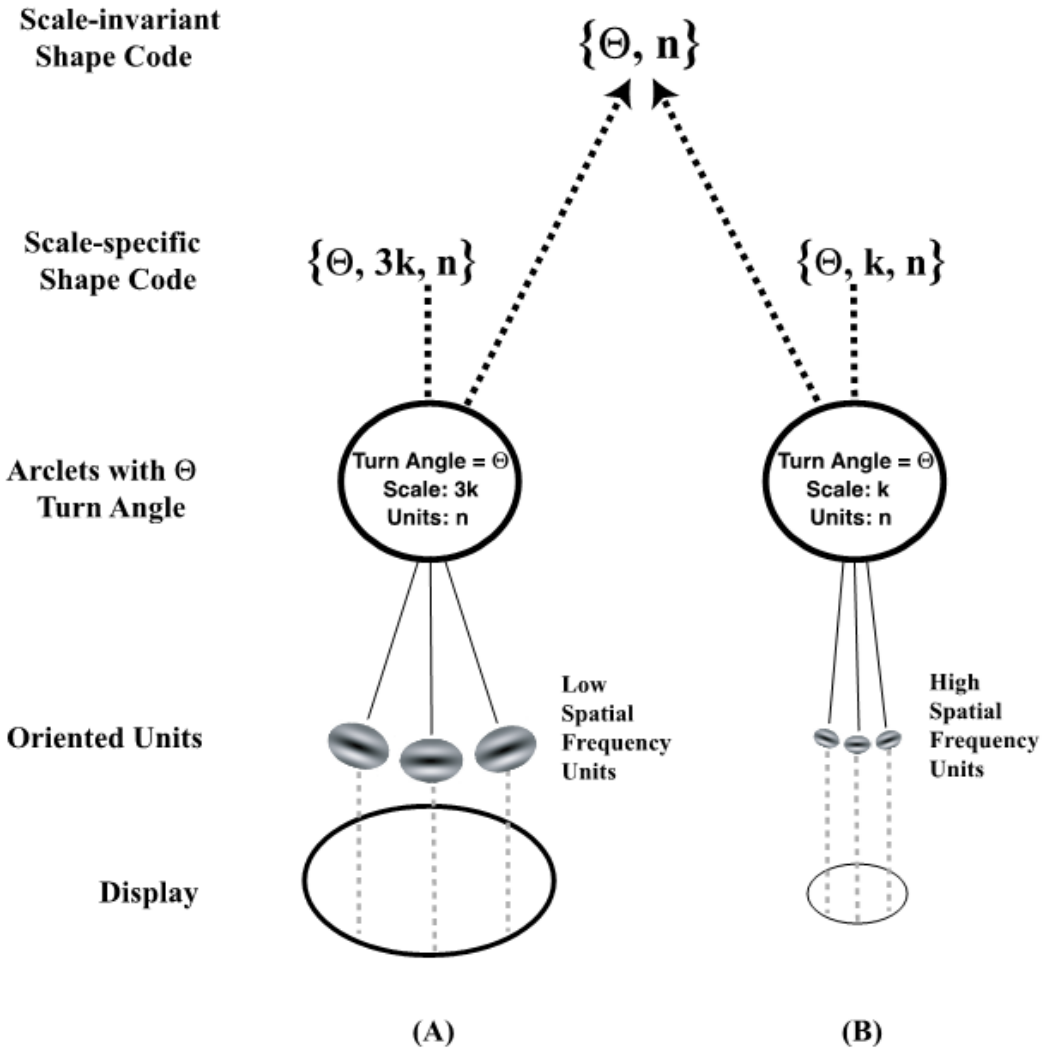

Figure 12. Scale invariant and scale specific shape coding from arclets. A) Large ellipse display. A contour fragment of approximately constant curvature initially activates oriented units along its boundary. *Arclet* units -- pairs of coaxial oriented units related by a constant turning angle – are activated along the contour segment. A constant curvature segment is encoded based on the extent of the boundary that fits arclets of a constant turn angle. The curve may activate arclets at several scales, but a crucial description is given by

the *largest* scale arclets that fit the segment adequately (below some threshold error). Here, the largest adequately fitting arclets for the segment are given as scale 3k. The turn angle along this segment is Θ, and the extent of the constant curvature segment best responded to by arclets of this turn angle and scale is n units. B) Small ellipse display. Arclet units are activated by small coaxial oriented units along the corresponding contour segment as in (A) that are related by a constant turn angle Θ. Here, the largest adequately fitting arclet has scale k, turn angle Θ, and extent of n units. A scale-specific representation of this segment of the ellipse is given by the three parameters of the fitting arclets: turn angle, scale, and number of oriented units comprising the extent of the segment. This scale-specific representation will differ for the large and small ellipse. Encoding at the largest adequately fitting scale makes available a scale-invariant representation. Omitting the scale parameter, the two segments of the two ellipses have the same shape because they match on the two parameters of turn angle and number of segments (of the largest adequately-fitting arclets in each case). The visual system's use of isotropic operators at different scales and the detection of curvature from sets of straight oriented units related by constant turn angles makes this scale invariant code available without special computation (e.g., normalization). Other parts of the contour will have other best-fitting arclets to signal approximately constant curvature segments. The complete contour representation of each ellipse consists of several joined segments of approximately constant curvature.

In the detector model, we have proposed that the visual system extracts smoothly curving contours from a set of discrete straight segments related by a turn angle. The visual system's earliest detectors are tuned to (and signal) straight oriented contrasts rather than curvature, even though ecological work on scene statistics suggests that most object contours are curved (Sigman, Cecchi & Magnasco, 2001; Chow, Jin & Treves, 2002). The multiscale representation scheme we have described is intriguing because it suggests a benefit for describing curves with small straight segments. If detectors in early visual areas were sensitive to true curvature rather than orientations of straight segments, then it would be much harder for the visual system to find equivalence between shapes whose absolute curvature differ due only to a change in scale. With straight segments, the curvature and extent of a contour will match across scales as long as the scale of the arclet used to describe the contour varies with the overall scale of the contour.

Symbolic Description

The final step in shape representation is to form a symbolic code of the shape. In this abstract description, the subsymbolic activations of oriented units are described by a small number of numerical variables. We propose that a shape's code is a description of the constant curvature segments in the shape, which include curvature, orientation, position, angular extent, and scale. Below, we briefly describe how each of these variables are obtained from the detector model.

Curvature: A segment's curvature is simply encoded as the turn angle between oriented units in that region of the contour. Turn angle at the finest scale is computed in the segmentation step. If the segment's curvature is low enough to be represented by larger scale detectors, turn angle increases proportionally with change in scale of the detector.

Orientation: A single orientation value of a detector is enough to describe a constant curvature segment. Arbitrarily, the visual system's symbolic code for a shape could include the orientation of the first unit in the counterclockwise direction of the segment. Then, the orientation of all other units will be fixed by their distance (i.e., the number of units) from the first unit and the curvature of the segment. Constancy across rotation can be obtained by computing segments' orientations relative to each other. Then, if the whole object is rotated, relative orientations are preserved.

Position: As for orientation, the position of a segment can be encoded as the spatial position of a single detector in the segment. The positions of all other detectors can then be determined by the curvature of the segment and the number and size of detectors between them. A more object-centric code of relative position can be computed as the normalized distance between oriented units in different segments.

Angular Extent: The angular extent of the segment is determined by its curvature and the number of oriented units in the segment. Since curvature is already a separate variable, angular extent can simply be encoded as the number of units in the segment.

Scale: The scale of detectors depends on the absolute curvature of the contour region, as described in the section about multiscale representation. The visual system describes a segment with the largest detectors whose error is still below some threshold. The lower the curvature of a contour region, the larger the detectors the visual system will use to describe it. Consequently, two shapes that differ by a factor of scale will have different scale variables, but all other variables will match. Ignoring or computing the relative scale of segments gives the symbolic description size invariance.

Conclusion and Future Directions

In this chapter, we have sketched a model that bridges the gap from subsymbolic activations of units sensitive to light contrasts to a fully symbolic description of an object's shape. First, it estimates curvature at the finest scale through a population code of arclets broadly tuned to various turn angles. The model then partitions the contour into segments of curvature, aiming to reduce the curvature variance within a segment. Next, it assigns a scale of detectors in each segment based on the curvature estimated by detectors at the smallest scale. Finally, a symbolic code for the shape is formed that describes each segment and relations between segments by a small number of numerical variables.

Though stages of the model have individually shown promising results, some work remains to be done developing a fully connected pipeline. First, we would like to reduce noise in the curvature profile estimated by population coding and use that as input to the segmentation algorithm instead of the contour's true curvature. Experimental work also remains to be done

finding the visual system's tolerance for error in describing curved regions of a contour with larger scale detectors. Once this error tolerance has been parameterized, we can model the largest-good-enough scale at which a segment is described by arclets and get a fully symbolic code for the shape.

While interesting problems remain to be solved, initial results for the detector model appear promising. The model is more biologically plausible than the computational model we previously proposed, as it is built up from oriented detectors and suggests a way of computing curvature subsymbolically instead of through computations unlikely to exist in neurophysiology. The detector model also offers a solution to the puzzle of locality and scale invariance that emerged in the computational model. Constant curvature segmentation is done based only on local curvature differences, but the use of relative curvature and multiscale representations offer equivalence between shapes that differ only in size. Finally, the detector model can be parameterized to have outputs that closely align with outputs of the computational model, which is beneficial because these outputs were found to account for unique predictions in perceptual performance.

References

Baker, N., & Kellman, P. J. (2018). Abstract shape representation in human visual perception. *Journal of Experimental Psychology: General*, *147*(9), 1295.

Baker, N., Garrigan, P., & Kellman, P.J. (2020). Constant Curvature Segments as Building Blocks for 2D Shape Representation. Under revision for *Journal of Experimental Psychology: General.*

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, *24*(4), 509-522.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 121.

Celebrini, S., Thorpe, S., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake primate. *Visual neuroscience*, *10*(5), 811-825.

Chen, Y., Geisler, W. S., & Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature neuroscience*, *9*(11), 1412-1420.

Chow, C. C., Jin, D. Z., & Treves, A. (2002). Is the world full of circles?. *Journal of vision*, *2*(8), 4-4.

Clifford, C. W., Wyatt, A. M., Arnold, D. H., Smith, S. T., & Wenderoth, P. (2001). Orthogonal adaptation improves orientation discrimination. *Vision research*, *41*(2), 151-159.

Cooper, L. A., Schacter, D. L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 43.

Gibson, J. J. (1966). The senses considered as perceptual systems.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA, US.

Graf, A. B., Kohn, A., Jazayeri, M., & Movshon, J. A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature neuroscience*, *14*(2), 239.

Heitger, F., Rosenthaler, L., Von Der Heydt, R., Peterhans, E., & Kübler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision research*, *32*(5), 963-981.

Hickinbotham, S. J., Hancock, E. R., & Austin, J. (1997). S-Gabor channel design for segmentation of modulated textures.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, *195*(1), 215-243.

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of neurophysiology*, *73*(1), 218-226.

Kellman, P. J., & Garrigan, P. (2007). Segmentation, grouping, and shape: Some Hochbergian questions.

Lüdtke, N., Wilson, R. C., & Hancock, E. R. (2000, May). Tangent fields from population coding. In *International Workshop on Biologically Motivated Computer Vision* (pp. 584-593). Springer, Berlin, Heidelberg.

Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex*, *4*(5), 523-531.

Ma, W. J. (2010). Signal detection theory, uncertainty, and Poisson-like population codes. *Vision research*, *50*(22), 2308-2319.

Marr, D. (1982), *Vision: A Computational Approach*, San Francisco, Freeman & Co.

Pei, X., Vidyasagar, T. R., Volgushev, M., & Creutzfeldt, O. D. (1994). Receptive field analysis and orientation selectivity of postsynaptic potentials of simple cells in cat visual cortex. *Journal of Neuroscience*, *14*(11), 7130-7140.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of neurophysiology*, *88*(1), 455-463.

Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, *387*(6630), 281-284.

Schmidtmann, G., Jennings, B. J., & Kingdom, F. A. (2015). Shape recognition: convexities, concavities and things in between. *Scientific reports*, *5*, 17142.

Sigman, M., Cecchi, G. A., Gilbert, C. D., & Magnasco, M. O. (2001). On a common circle: Natural scenes and Gestalt rules. *Proceedings of the National Academy of Sciences*, *98*(4), 1935-1940.

Smits, J. T., & Vos, P. G. (1987). The perception of continuous curves in dot stimuli. *Perception*, *16*(1), 121-131.

Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2018). High-capacity preconscious processing in concurrent groupings of colored dots. *Proceedings of the National Academy of Sciences*, *115*(52), E12153-E12162.

Chapter 3: Shape Perception from Unconnected Elements: Smoothness Constraints Revealed by Convergent Measures

Nicholas Baker[1*] & Philip J. Kellman[1]

1. Department of Psychology, University of California, Los Angeles

*Corresponding author. Please direct any questions to nbaker9@ucla.edu

Abstract

A remarkable phenomenon in perception is that the visual system spontaneously organizes sets of discrete, spatially separated elements into abstract shape representations. We studied perceptual performance with dot displays in various tasks to discover what spatial relationships support contour and shape perception. In Experiment 1, we tested conditions that lead dot arrays to be perceived as smooth contours vs. having vertices. In preliminary work, we found that a dot triplet is perceived as a corner if virtual lines connecting the dots intersected at ninety degrees or less. We generated dot arrays with and without such "vertex triplets" and compared participants' phenomenological reports of shape for dot arrays that were sampled from shapes with smooth curves vs. shapes with angular corners. People tended to give higher shape ratings for dot arrays from curvilinear shapes. In Experiment 2 we tested shape encoding using a mental rotation task (Shepard & Metzler, 1971). On each trial, subjects judged whether two dot arrays were the same or different. RT was measured at five angular differences. Subjects responded faster for displays without vertex triplets, suggesting economical encoding of smooth displays. We followed up this result in Experiment 3 using a visual search task. Shapes with and without vertex triplets were embedded in arrays with 25 distractor dots. Subjects were asked to detect which display in a 2IFC paradigm contained a shape against a distractor with random dots. Performance was better when the dots were sampled from a smooth shape than when they were sampled from a shape with vertex triplets. These results suggest that the visual system processes dot arrangements as coherent shapes automatically using precise smoothness constraints. This ability may be a consequence of processes that extract curvature in defining contour and object shape.

One of the primary functions of the visual system is the formation of abstract shape representations. In natural vision, encoding a representation of shape frequently involves organizing spatially distinct elements into a configural whole, as objects tend to occlude each other. Classic work by the Gestalt psychologists has shown that the whole that we encode from these fragments is fundamentally different from the sum of individual parts (e.g., Koffka, 1935; Lindemann, 1922). Wertheimer (1923) identified several stimulus features by which distinct elements could be organized together, such as similarity, proximity, common motion, and good continuation. Related work has also studied how we interpolate contours and surfaces behind occluders based on geometric and surface properties of the fragments (Michotte, 1964; Kellman & Shipley, 1991; Yin, Kellman & Shipley, 1997) and how we integrate paths between disconnected oriented elements (Field, Hayes & Hess, 1993).

One phenomenon that these research efforts do not explain, however, is how the visual system organizes an array of dots like the one shown in Figure 1a into a single shape description. The shape is unfamiliar, so no template matching can explain the percept. Nor can similarity in the elements' size, color, or individual shapes account for their organization into a configural whole, as shown in Figure 1b-d. Moreover, unlike the displays used in research on path integration, the dot elements have no explicit orientation or tangent direction. The visual system could, in principle, interpolate any number of possible contours between this dot array. Despite these difficulties, research has shown that dot displays give rise to shape representations that are invariant to changes in size, orientation, and position (Baker & Kellman, 2018). There also seems to be a large degree of consistency across shape representations formed by different observers. We found that subjects had no ability to detect changes in dot positions when dots were moved along a never-before-seen virtual contour from which the first array of dots was

sampled. This suggests that the shape representation viewers form in their mind very closely resembles the physical contour from which the dots were originally sampled (see Figure 2).

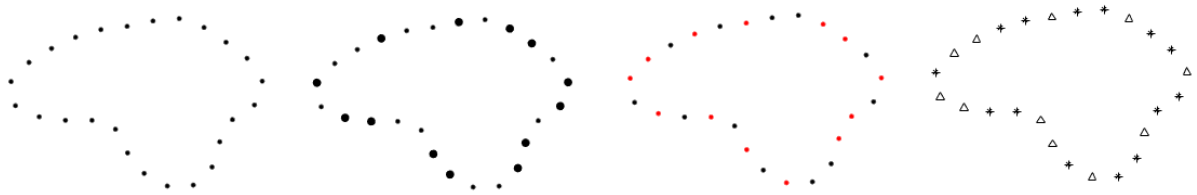

Figure 1. Different dot displays for the same shape. We organize the elements in each of these four displays into a unified shape representation despite differences in element size, color, and shape.

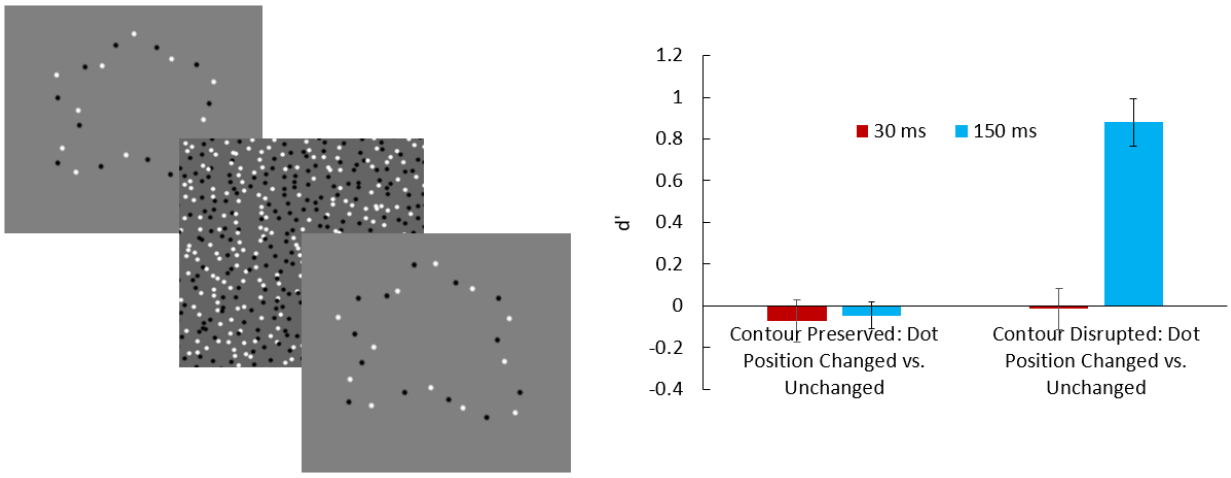

Figure 2. A: Trial from Baker et al. (2018) in which dots are shifted along a shape's virtual contour. B: Sensitivity to shape change. When dots were shifted along the virtual contour, participants had no sensitivity to the change, even with 150ms of exposure time.

As remarkable as it is that the visual system encodes shape representations from unconnected dots, not all dot arrays give rise to a shape percept. Studies on the percept of shape from randomly placed dots have found that certain configurations are much more frequently perceived to have shape than others, depending on proximity and good continuation between the dot elements (van den Berg, 2006). Considerable research has been done on how the visual system organizes an array of dots into multiple distinct shapes, or into a single shape among

random dot distractors. Proximity appears to play a major role in these computations (van Oeffelen & Vos, 1983; Kubovy and Wagemans, 1995; Papari & Petkov, 2005), although similarity (Zucker, Stevens & Sander, 1983) and good continuation are also important cues (Smits, Vos & van Oeffelen, 1985; Lezama, Randall, Morel & van Gioi, 2016).

Understanding what spatial relationships between dots influence whether an array is perceived as a shape could provide crucial insight into how the visual system perceives shapes in general. One important question that dot arrays could help answer is how the visual system encodes contour information. In typical shape displays, the contour is physically present in the stimulus, so it can be difficult to determine how the visual system abstracts and encodes the shape in visual memory. In dot arrays, no contour is physically present, so they provide unique insight into what contour features the visual system naturally imputes to a shape representation.

One way the visual system could form a contour representation from unconnected dots is by interpolating straight edges between adjacent boundary points (e.g., O'Callaghan, 1973). This process would be computationally simple and parsimonious in the sense that the contour representation is highly constrained by the set of points in the proximal stimulus. On the other hand, we have the phenomenological experience of perceiving smooth curves in displays like the dots in Figure 1. Moreover, connecting straight edges at each point is equivalent to encoding the spatial position of each non-colinear dot. Corners are perceptually important features of a contour (Attneave, 1954), so shifting dots along the contour would produce a very different shape representation if each dot was first order discontinuous corner. Past results have shown that this kind of shift is not detectable to human observers (Baker & Kellman, 2018).

There is also empirical evidence that the visual system does not always represent dots as corners in a contour. Koffka (1931) placed points along a circle to estimate the number of dots at

which the virtual contour was perceived as smooth. He estimated that this transition occurred at around eight evenly spaced dots (i.e., an inclusive angle of 135 degrees). Bouma (1976) gave a more conservative estimate that 10 dots were needed for the virtual contour to be perceived as smooth. Smits and Vos (1987) used more systematic tests to estimate this transition point and found that when the angle between triplets of dots was sufficiently large (greater than about 140 degrees), the dots were perceived as curvilinear (Figure 3a). In a later study, van Assen and Vos (1999) developed a more objective measure of perceived curvilinearity by measuring participants' bias to say whether a target dot was below or above a virtual contour defined by four other dots. They found that when the inclusive angle between the central dots was 135 of 150 degrees, participants' bias was consistent with perceiving a curved contour. Feldman (1996) systematically varied the inclusive angle for three dot displays and found that the 50% threshold for curvilinear responses vs. angular responses was at around 120 degrees. He found that when a fourth dot was added to the configuration to create two similar angles from dot triplets, the threshold went down (Feldman, 1997) (Figure 3b). In a similar experiment, we created five-dot displays in which the sign of curvature for the triplets with the second and fourth dots as their vertex was the same or different than the sign of curvature for the triplet with the third dot as its vertex (Figure 3c-d). We found that the threshold for curvilinearity was much higher when the flanking vertices had the same sign of curvature as the central vertex than when their curvature was of opposite sign, but that there is a hard threshold at 90 degrees regardless of the geometry of other dots (Baker & Kellman, unpublished).

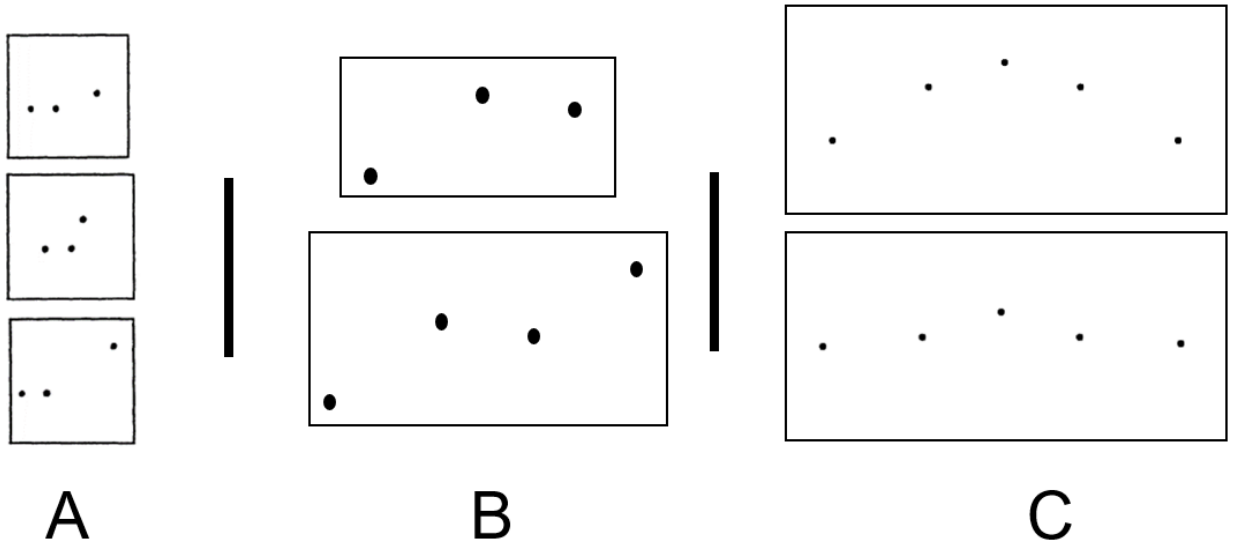

Figure 3. A: Dot triplets used by Smits and Vos (1987), reprinted from the original article. B: Comparison of a dot triplet and quadruplet from Feldman (1997). Feldman found that the addition of a fourth dot with similar turn angle increased the probability of a curvilinear response. **C: Arrangement of five dots from Baker & Kellman (unpublished).** Despite the two arrays having an identical configuration of the middle three dots, the percept of a corner at the center dot is affected by the position of the first and fifth dots.

If forming a shape representation by encoding each dot position as a corner is both parsimonious and computationally simple, why would the visual system ever extract contour representations contour representations with smooth curves from arrays of dot arrays? One possibility is that shape contours with smooth curves are *representationally* simpler even if they are more complicated to compute. Past research on connected contours has shown that perceptual tasks requiring a shape representation are accomplished better and more quickly with smooth contours than with angular contours (Bertamini, Palumbo & Redies, 2019). Other work has shown that visual system has special facility for encoding contours of constant curvature (Baker, Garrigan & Kellman, under review). A perceptual corner has a first order discontinuity at its vertex, meaning there will always be a part boundary at that point. On the other hand, if the contour is perceptually smooth at that point, the entire segment could be represented as a single part. As a consequence, though there may be a greater upfront computational cost to

interpolating smooth curves between points in a dot array, the resulting shape representation will ultimately be simpler, as it is made up of fewer parts. Relatedly, the presence of a corner signals a L-junction, a common cue for intersecting contours belonging to different shapes in visual perception (Shipley & Kellman, 1990). L-junctions might make it more difficult to perceive the contour as naturally continuing at the corner point, forcing the visual system to assess if the junction is formed by one or two shape boundaries. All of this predicts that the abstract shape representation that is ultimately encoded by an array of dots might be stored more efficiently as a set of relatively few curved segments than a larger set of straight segments, even if this requires more initial processing to compute. If this is the case, we would expect that it will be easier to encode arrangements of dots that appear to have few or no sharp corners as a shape than arrangements of dots with many perceived contours.

We tested whether dot arrays whose spatial relationships signaled smooth curves were more or less easily encoded as shapes using both subjective and objective measures. In Experiment 1, we sampled dots from shapes with smooth curves and from shapes with sharp corners and asked participants to judge how much each dot array looked like a shape. In Experiment 2, we showed pairs of dot arrays sampled from smooth or corner shapes at different orientations and measured the time it took for participants to judge whether the shape formed by the dots was the same or different. In Experiment 3, we hid dot arrays sampled from both kinds of shapes among randomly placed distractor dots and measured participants' ability to detect the shape in both conditions.

Experiment 1

The goal of Experiment 1 was to measure the subjective strength of a shape percept in three different kinds of dot displays. We created arrays of dots by sampling novel shapes with

smooth curves, by sampling novel shapes with sharp corners, and by randomly arranging them. We then asked subjects to rate the degree to which the dots appeared to form a shape outline. Our prediction was that dots sampled from smooth contours would be judged more shape-like than dots sampled from shapes with corners, which would in turn be judged more shape-like than random dot arrangements.

Method

Participants

Twenty-five undergraduates (3 male, 22 female, *M*age = 20.6) from the University of California, Los Angeles participated in the study for course credit. All participants had normal or corrected-to-normal vision.

Display and Apparatus

Subjects were seated 70 cm from a 20-inch View Sonic Graphic Series G225f monitor. The monitor was set to 1024x768 resolution, with a refresh rate of 100 hz. Except when noted otherwise, all aspects of the displays and apparatus in subsequent experiments were the same as in Experiment 1.

Stimuli

Experiment 1 included three different kinds of dot arrays: smooth, corner, and random. Smooth dot arrays were created by generating a smooth contour, which we formed by moving 12 control points toward or away from the center of a circle a random distance, then fitting cubic splines through the 12 control points in polar space (see Figure 4a). We then sampled 25 points along the contour to get the dot array. The points were sampled nonuniformly by taking 25 evenly sampled points and moving them in a random direction along the contour. Though not directly relevant to this experiment, we included jittering along the contour to prevent

participants from using local spatial relationships between a small set of dots rather than the overall shape of a dot array in future objective experiments (see Experiment 2 for more explanation). The distance points were moved from even spacing the contour was sampled from a normal distribution, with a mean distance of zero and a standard deviation equal to the contour length divided by 200. We constrained the display to never have two points fewer than 5 pixels away from each other (Figure 4d).

The corner dot arrays were created by generating a smooth dot array and then changing the spatial relationships between dots so that they formed perceptual corners (i.e., the angle between them was 90 degrees or less). For a given dot triplet, ABC, we imposed a corner percept by interpolating a line between points A and C, then moving point B perpendicularly away from the interpolated line while simultaneously moving A and C along the line until the vertex at B was between 78 and 90 degrees. This range was chosen so that the angular shapes would be reliably perceived as corners while also including some natural variability in the angle between dot triplets. We then interpolated straight lines between the 25 repositioned points to get a shape contour with corners and sampled 25 new dots using the same nonuniform sampling procedure used for smooth dot displays to get 25 dots (Figure 4b and 4d).

The random dot condition also began with the smooth dot array. Rather than moving dots to reduce the angle between them, dots were moved in a random direction. Each of the 25 points was moved a distance equal to the total length of the contour divided by 25 in a random direction, with the constraint that no two dots could be fewer than 5 pixels apart (Figure 4e). We used this method instead of truly random placement to prevent subjects from judging shape based on whether there was an open center within the dot array. The random condition also had a center with few or no dots, but we still expected it would not be perceived as a shape.

Dot arrays for each of the three conditions were matched in size, subtending on average 12.0 degrees of visual field, and at most 20.7 degrees of visual field.

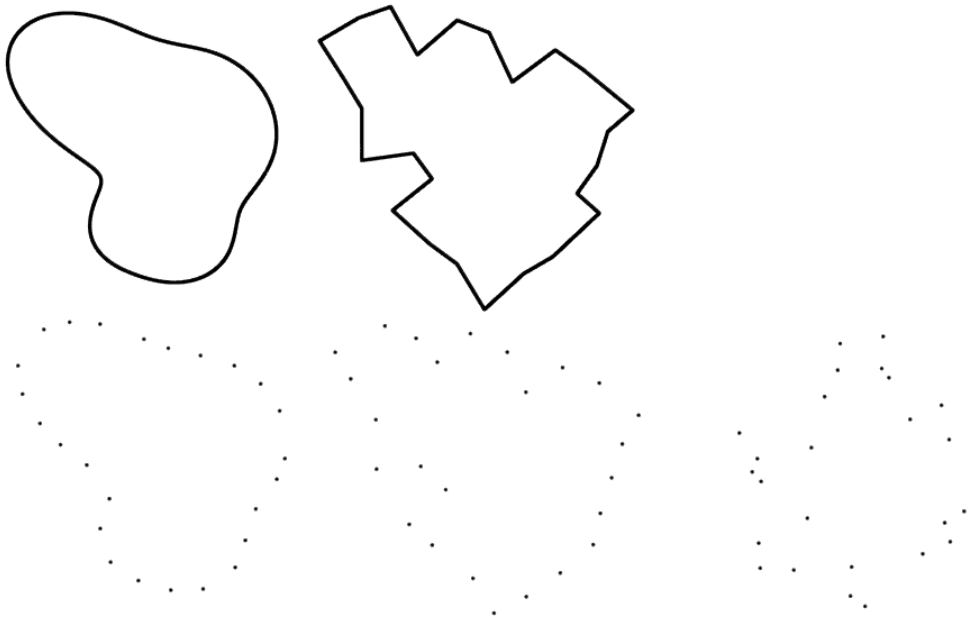

Figure 4. Shape stimuli used in Experiment 1. Top: A smooth and angular shape. Bottom (left to right): Dots sampled from the smooth shape, dots sampled from the angular shape, and random dot arrays.

Design

Experiment 1 had three conditions, with 70 trials in each condition. In the first condition, we showed dot arrays from smooth shapes. In the second condition, we showed dot arrays from shapes with corners. In the last condition, we showed random dot arrays. Trials with all three conditions were randomly interleaved.

Procedure

On each trial, participants were shown one of the three stimulus types and asked to evaluate the degree to which the dot array seemed to form a shape outline. The dot array remained on the screen until a response was given. Participants were instructed to rate the display on a 6-point scale, ranging from “The dots look totally random” to “The dots look totally

like a shape". By choosing an even number of points on the scale, there no neutral condition. Subjects always had to say if the dots looked more shape-like (ratings greater than 3) or more random (ratings 3 or less). We instructed participants to use the full range of numbers to reflect qualitative differences in the degree to which different dot arrays appeared to be shapes. Before beginning the main experiment, participants completed five practice trials to familiarize themselves with the response buttons and to expose them to all three stimulus types before giving recorded shape judgments. A sample trial is shown in Figure 5.

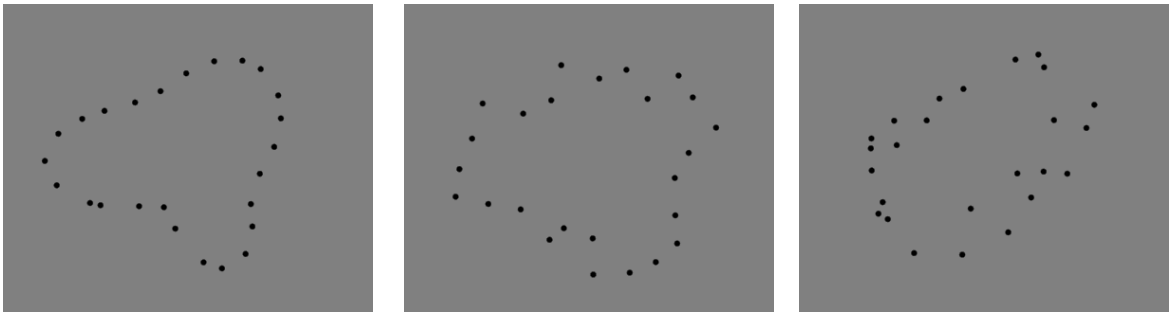

Figure 5. A sample trial for each of the three conditions in Experiment 1. A: Trial with a curvilinear shape. B: Trial with an angular shape. C: Trial with randomly positioned dots.

Results

The mean subjective rating for each of the three stimulus types is shown in Figure 6a. There is a clear ordering in which dots sampled from smooth contours are perceived as most shape-like, followed by dots sampled from angular contours, followed by randomly sampled dots. This pattern was reflected for every participant who completed the experiment. A one-way ANOVA confirmed a significant difference between the groups, $F(2, 57) = 181.05$, $p < .001$, and Bonferroni corrected paired sample t-tests confirmed that dots sampled from smooth contours were rated more shape-like than dots sampled from angular contours, $t(19) = 13.22$, $p < .001$, and that dots sampled from angular contours were rated more shape-like than randomly sampled dots, $t(19) = 21.04$, $p < .001$.

We also analyzed the average number of trials in which subjects perceived a shape at all. For this measure, we included any display that received a subjective rating greater than 3. The results are shown in Figure 6b. Paired samples t-tests confirmed that subjects' perceived significantly more of the dots sampled from smooth contours as a shape than they did dots sampled from angular contours, $t(19) = 5.60$, $p < .001$ and that dots sampled from angular contours were perceived as shapes significantly more often than randomly sampled dots, $t(19) = 12.65$, $p < .001$.

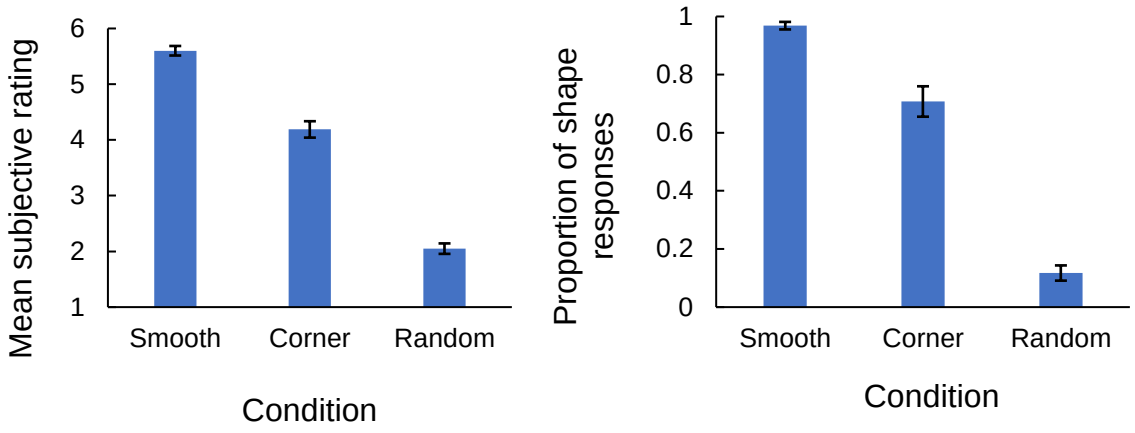

Figure 6. Experiment 1 results. A: Participants rating of shape for each of the three conditions. B: The percentage of ratings that were more shape-like (i.e., rating > 3) for each condition.

Discussion

Experiment 1 gave evidence that dots sampled from smooth contours are more phenomenologically shape-like than dots sampled from contours with sharp corners. Every participant gave higher shape ratings for the smooth contour condition and reported perceiving more of the smooth contours as shapes than the angular contours. These results also support the notion that dots sampled from smooth shapes are a qualitatively different kind than dots sampled from angular shapes. Participants never saw the underlying contour from which the dot arrays were sampled. Geometrically, all the dots sampled from shapes with corners could be represented with curvilinear contours. If subjects perceived the dots sampled from angular

contours as smooth, we would expect very little phenomenological difference between the two non-random dot conditions. Instead, we find that dots sampled from angular contours are consistently perceived differently than the dots sampled from smooth contours.

Experiment 1 provided a subjective measure of a dot array's perceived shape-ness but could be influenced by biases or demand characteristics. Still, it is important to know what subjects believe they are seeing, and subjective reports give a direct answer to this phenomenological question. We assess differences between the perceived shape-ness of dots sampled from smooth and angular contours with objective measures in Experiments 2 and 3.

Experiment 2

One of the key functions that encoding an abstract shape representation serves is allowing comparison of shapes across different orientations (Baker & Kellman, 2018). In Experiment 2, we compared subjects' ability to encode a shape representation for dots sampled from smooth and angular contours by testing them on a shape matching mental rotation task. Inspired by Shepard and Metzler (1971), we simultaneously presented two differently oriented dot arrays and asked subjects to judge whether they defined the same virtual contour. We expected that if dots sampled from smooth contours are more naturally perceived as shapes, subjects should have an advantage in the mental rotation task.

Method

Participants

Participants included 25 undergraduates (4 male, 21 female, *M*age = 19.8) from the University of California, Los Angeles who enrolled in the study for course credit. All participants had normal or corrected-to-normal vision.

Stimuli

Smooth and angular dot arrays were generated as in Experiment 1. In Experiment 2, each array was a member of a pair with either the same shape or a different shape. When the shape was the same, we used the same virtual contour, but sampled a different set of dots so that local spatial relations between dots could not be used as a cue. When the shapes were different, we generated the second member of the pair by moving one of the control points for the original shape a random distance between 1.93 and 4.11 degrees toward or away from the center of the shape. We then randomly selected an adjacent control point to the one we just moved and moved it toward or away from the center such that the total contour length for the new shape was the same as the total contour length for the original shape (see Figure 7 for a pair). For angular shapes, we then applied the same set of changes described in Experiment 1 to the new shape. Dot arrays also differed in orientation. In each trial, the second dot array could be rotated 0, 45, 90, 135, or 180 degrees relative to the first.

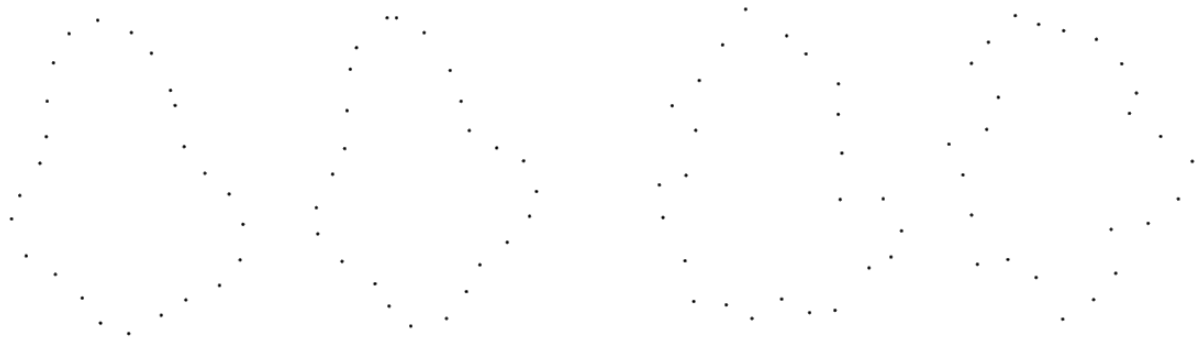

Figure 7. Pairs of smooth and angular shapes used in Experiment 2.

Design

The experiment consisted of 200 trials, half of which showed shape pairs sampled from smooth virtual contours, and half of which showed shape pairs sampled from angular contours.

For each of these two conditions, there were 20 trials at each of the five angular differences, 10 of which included the same shape, and 10 of which included different shapes.

Procedure

On each trial, two arrays of dots were shown on the screen simultaneously, one centered in the left half of the monitor screen, and one centered in the right half. Subjects were instructed to look at both dot arrays and determine whether the shape defined by each array of dots was the same or different, irrespective of a difference in orientation and the local positions of dots. The two dot arrays remained on the screen until subjects responded. Participants were told that response time was being measured, but that they should emphasize responding correctly over responding quickly. Before beginning the main experiment, subjects completed 12 practice trials to familiarize themselves with the task. Performance in the practice trials was not analyzed. A sample trial for each condition is shown in Figure 8.

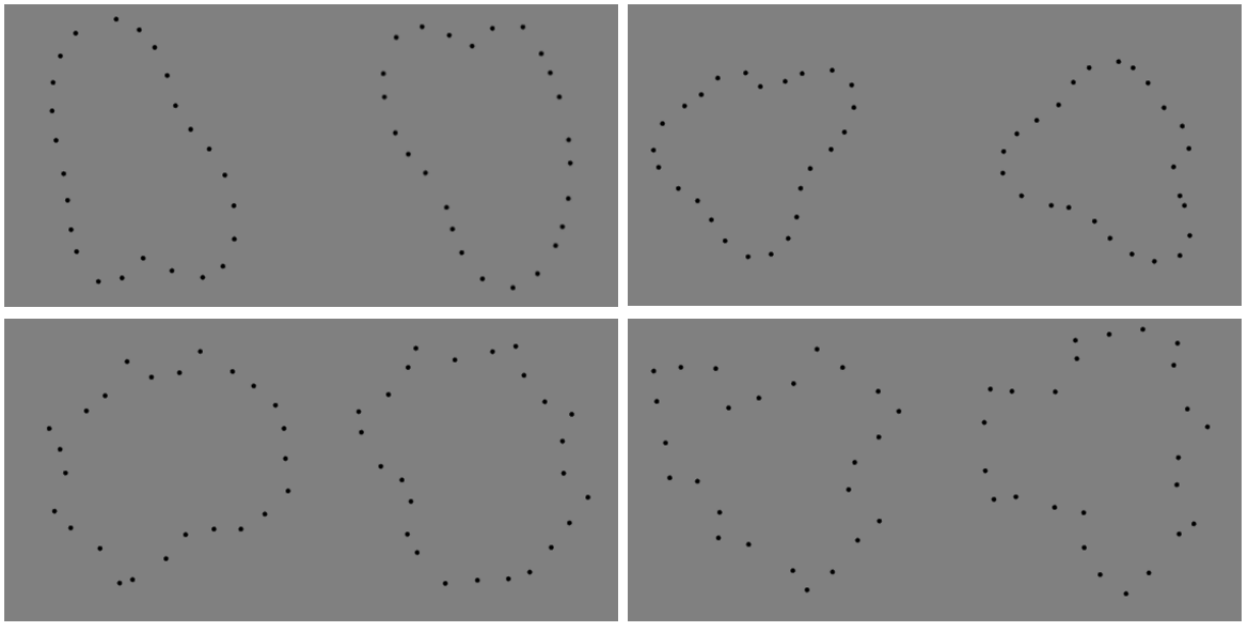

Figure 8. Sample trials from Experiment 2. Top left: Dots sampled from smooth contours with the same shape. Bottom left: Dots sampled from angular contours with the same shape. Top right: Dots sampled from smooth contours with different shape. Bottom right: Dots sampled from angular contours with different shape.

Results

Following Shepard and Metzler (1971), we analyzed the reaction time only for trials in which the two shapes were the same and subjects responded correctly. Mean response times for each angular difference are shown in Figure 9a. A repeated measures ANOVA confirmed a significant main effect for the type of shape from which the dots were sampled, $F(1,25) = 5.33$, $p = .03$. A linear regression test found no effect of angular difference on reaction time, either on the dots sampled from smooth contours, $F(1, 128) = 0.39$, $p = .54$, or on angular contours, $F(1,128) = 0.48$, $p = .49$.

Superiority in performance for dots sampled from smooth contours was also reflected in accuracy measures (Figure 9b). ANOVA analyses confirmed that accuracy was significantly higher in displays in which the dots were sampled from smooth contours than when they were sampled from contours with sharp contours $F(1,25) = 40.92$, $p < .001$.

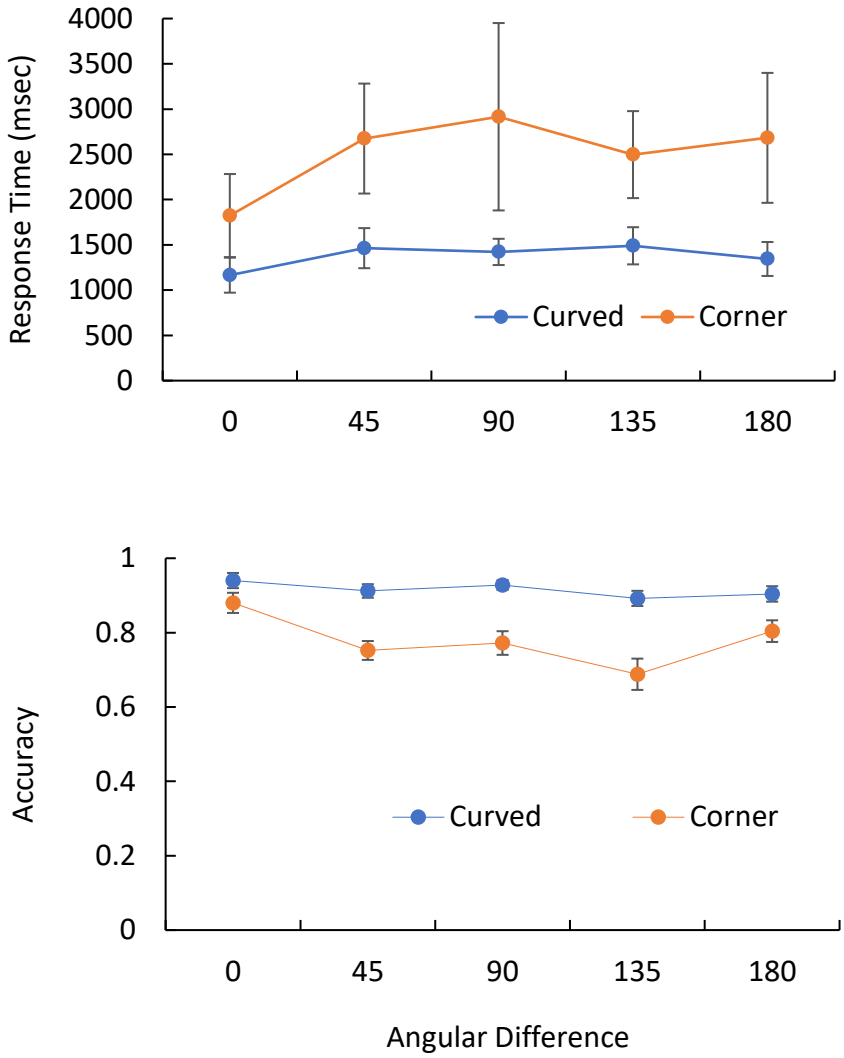

Figure 9. Response time and accuracy for Experiment 2. Top: Response time on correct trials as a function of angular difference. Bottom: Accuracy as a function of angular difference.

Discussion

Experiment 1 found that participants rated dots sampled from contours with sharp contours as less shape-like than dots sampled from smooth curves. In Experiment 2, we tested whether these subjective differences would be reflected in perceptual performance. Because they were rotated and had positions along the contour resampled, the target pairs of dot arrays we showed in Experiment 2 differed from each other both in absolute orientation and in terms of the

specific positions of the elements with respect to each other. Accurate responding for the task therefore required forming a representation of a shape's contour from the set of dots that was object-centric and invariant to orientation changes. Differences in response time and/or accuracy for the two kinds of dot arrays therefore presumably correspond to the ease with which participants encoded the array as a shape.

We found that dots sampled from shapes with smooth contours could be compared across orientation changes more quickly than dots sampled from shapes with sharp corners, which suggests that these dot arrays are more easily encoded and perceived as orientation-invariant shapes than arrays sampled from shapes with sharp corners. Participants also made fewer errors when mentally rotating dots sampled from smooth contours than dots sampled from angular contours. Lower response times therefore cannot be explained by a speed-accuracy tradeoff.

One puzzling aspect of our data is that we found no reliable effect of magnitude of angular rotation on response time for either trial type. Shepard and Metzler's work (1971), from which we modeled our experiment, showed a strong linear relationship between response time and degree of angular difference for shapes rotated in the picture plane. Response time has also been shown to vary with degree of change from a canonical orientation in naming tasks for familiar objects (Jolicoeur, 1985). One possibility for why angular difference had such a small effect in our study is that subjects responded after a somewhat fixed period of time, even if more or less time was needed to make an accurate decision. This could explain why we see a reduction in accuracy as a function of angular difference in the sharp corner condition even though response time does not increase. Importantly, though, even if subjects are using a more fixed period of time, this amount of time is different for the smooth and sharp corner conditions. Participants consistently require more time to decide if dots sampled from a shape with sharp

corners is the same or different, even if response times do not increase monotonically with angular differences in either condition.

Another intriguing possibility is that dot configurations represent a special class of stimuli whose time for recognition does not scale with angular difference. Past work on mental rotation has shown that certain kinds of stimuli with salient landmark features have much flatter recognition slopes than stimuli without salient landmarks (Hochberg & Gellman, 1977). Flat slopes have also been found for familiar objects when participants were informed ahead of time what object they would be shown (Cooper & Shepard, 1973). Why mental rotation of dot patterns would have flat slopes is mysterious in view of these findings, as they are neither familiar nor do they have salient local features. In fact, any salient local feature obtained from a local group of dots in one of the arrays would not be present in the other matching array, since dot positions along the contour are independently sampled in matched pairs. One possibility is that the simplicity of dot arrays gives rise to flat mental rotation slopes. According to Hochberg and Gellman (1977), mental rotation of shapes will scale with angular distance if representations must be built up from successive glances. Possibly, the relatively few bits of information in an array of 25 dots can be extracted with only one glance. This is partially supported by previous findings that the spatial positions of an array of 25 dots are registered within the first 30 ms of exposure (Baker & Kellman, 2018).

.

Experiment 3

As we have discussed, the visual system has an amazing capacity for form contour representations from unconnected dots. In Experiment 3, we further tested these capabilities by showing dot displays embedded among a field of random noise dots. The experimental paradigm

was similar to one used by Uttal (1973) for dots along a curved or straight line segment. Uttal found that for these simple segments, participants had significantly more trouble detecting the target when it deviated more from a straight line, but there seemed to be little difference for angular vs. curvilinear deviations. In the present study, we tested participants' ability to detect whole forms defined by dots.

To do this, we used a two-interval forced choice (2IFC) paradigm in which one stimulus contained a shape embedded in noise and the other stimulus contained noise alone. Participants' task was to choose the interval that contained a coherent shape. In order to group together and describe the shape of a set of dots in noise, subjects would have to first use some spatial relationships between the dots in the array to identify which dots belonged to a shape outline and which were random. Critical cues like proximity could be potentially misleading for this kind of display. Manipulating the kind of shape contour that the target dots were sampled from, we tested participants' ability to decide which of the two intervals contained a shape and which consisted only of noise dots. We predicted that unlike in the case for simple segments, dot arrays sampled from whole shapes with smooth contours would be more easily detected than dots sampled from whole shapes with sharp corners.

Method

Participants

Twenty-six undergraduates (6 male, 20 female, *M*age = 21.6) from the University of California, Los Angeles participated in this study for course credit. All participants had normal or corrected-to-normal vision. One subject's data was excluded prior to analyzing his results because he did not appear to understand the instructions by the time he had finished the practice portion of the experiment.

Stimuli

Dot arrays from smooth and cornered shape contours were generated as in Experiment 1. In Experiment 3, however, the dots sampled from contours were hidden among 25 distractor dots. Distractor dots were created by uniformly sampling from the rectangle that circumscribed the target dots. Each trial also included a dot display with no shape. Rather than placing all 50 dots in the other display completely randomly, we created random displays of 25 dots as in Experiment 1, with the only difference being that we moved each dot twice the average distance between dots. This was to create displays with no shape that still had some emptiness in the middle of the array to prevent participants from using that as a low-level cue. We then added 25 dots by uniformly sampling from the circumscribing rectangle as in the target displays. Figure 10 shows a target display with dots from a smooth contour, a target display with dots from a corner contour, and a non-target display.

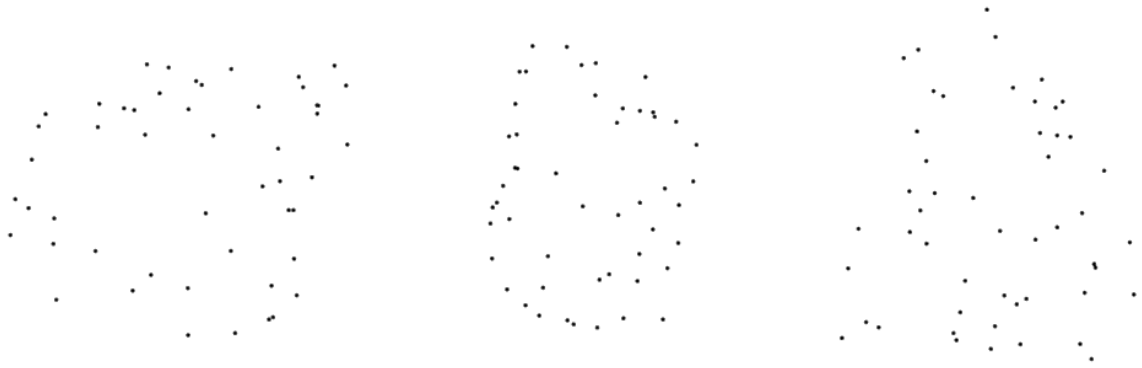

Figure 10. Target and distractor displays for Experiment 3. Left: Target display with dots sampled from an angular contour. Middle: Target display with dots sampled from a curvilinear contour. Right: Non-target display. In each trial, one of the two kinds of target display and the distractor display would be presented in a randomized order.

Design

The experiment had two conditions, one in which the target dots were sampled from a smooth contour and one in which the target dots were sampled from a corner contour. For both

conditions, half the trials had the target display shown first and half had the target display shown second. There were 120 total trials for each condition. Participants completed 12 practice trials before beginning the main experiment.

Procedure

Experiment 3 involved a 2IFC task in which one display consisted of dots sampled from a smooth or angular contour among noise dots and the other display consisted only of noise dots. Before beginning the experiment, subjects were told they would be looking for shapes hidden in dots. We showed 20 examples of shape displays made out of 25 dots, half of which were sampled from smooth contours, and half of which were sampled from corner contours. In each trial, we first presented a fixation cross at the center of the screen for 600 ms, then showed the first of the two dot displays for 800 ms. The dot display was then masked by a pattern of black and white dots for 500 ms, after which the second dot display was shown, also for 800 ms. This display was masked for 500 ms, and then subjects were asked to report whether a shape was hidden in the first or second of the two dot displays. Subjects were not cued to look for any specific shape in the displays and were told to pick whichever one they thought had dot arrangements that contained any shape. During practice, subjects were given feedback telling them if they were correct or incorrect and showing the hidden shape highlighted in white dots. Sample trial from the curved and angular condition are shown in Figure 11.

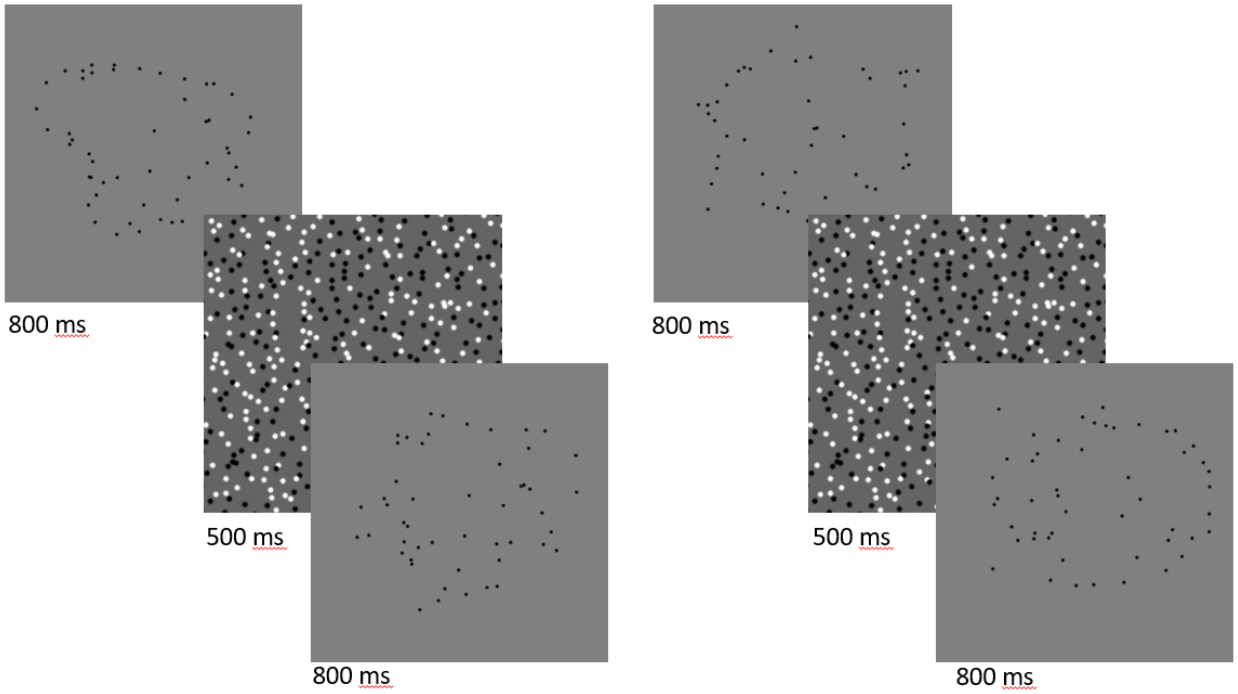

Figure 11. Sample trials from Experiment 3. Left: The target shape is sampled from a curved contour and is shown first. Right: The target shape is sampled from an angular contour and is shown second.

Results

The results for Experiment 3 are shown in Figure 12. Performance was significantly better than chance both when the dots were sampled from a smooth contour ($t(24)$ 17.59, $p < .001$) and when they were sampled from a corner contour ($t(24) = 15.69$, $p < .001$). Participants were significantly better at detecting the target dots when they were sampled from a smooth contour than when they were sampled from a corner contour ($t(24) = 10.3$, $p < .001$, Cohen's d = 1.1).

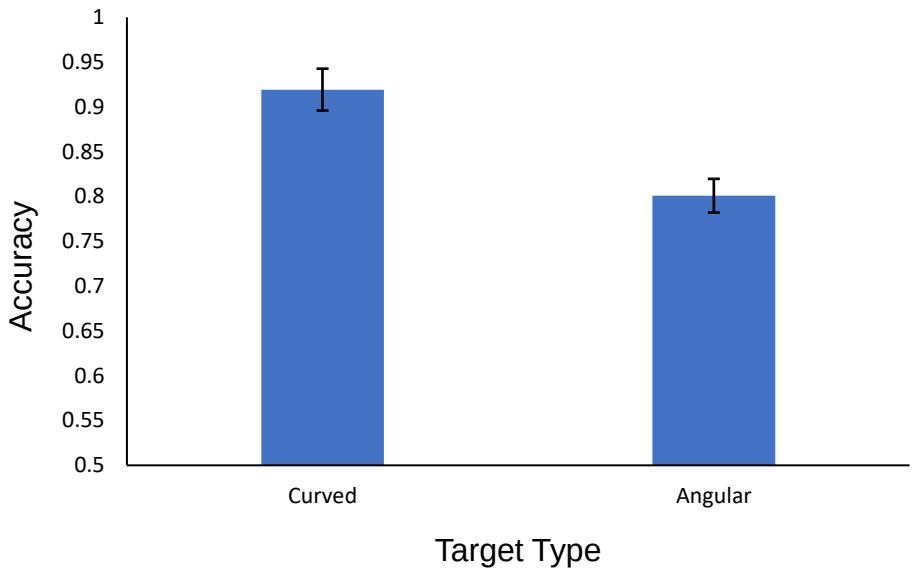

Figure 12. Accuracy for Experiment 3.

Discussion

In the target displays used in Experiment 3, the dot elements that formed the outline of a shape were physically identical to the distractor dots. The local spatial relationships between small groups of target dots are also not different from relationships between groups of distractor dots or groups that are a mix of targets and distractors. In this sense, the perceptual task subjects are performing is fundamentally different from the path detection task used by Field, Hayes and Hess (1993). In the classic path detection task, local orientation relationships between the dots determine whether the path is detected depending on the relatability of the local elements. The perceptual salience of paths likely depends on a contour-linking process that produces an intermediate representation in the process of contour interpolation (Kellman, Erlikhman & Kellman, 2016). Unlike the targets used by Uttal (1973), the local spatial relationships between target dots are not consistent. Neither the spacing nor the turn angle is the same between nearby dot triplets in our displays. In the Experiment 3 task, detection of the target depends on global

detection of a contour. The visual system might be considering multiple possible dot organizations and determining whether they configure into a global shape.

Participants' good performance for target dots sampled from both smooth and corner contours suggests that the visual system has a quite robust capability to detect a variety of shapes from distractors. However, there is also a clear performance advantage for detection of shapes with smooth contours over detection of angular shapes. One explanation for this difference might be that detection of a shape depends critically on object closure (Kovacs & Julesz, 1993). While dot arrays from both smooth and angular shapes are closed, the variance in turn angle between dot triplets in the smooth shape arrays are relatively small, while for the angular shape arrays the turn angle accentuated in a few areas (i.e., the corners). Possibly, the more consistent trajectories towards closure signal the presence of a target more strongly than a few large turns to achieve closure.

The observed difference is likely also a consequence of simpler representations for dots sampled from smooth shapes than those formed from dots sampled from shapes with sharp corners. If the visual system can represent the curvilinear shapes with a few structural primitives, but requires several to represent angular shapes, it follows that search for the more representationally complex targets would be less accurate. A similar effect and explanation have been given for search for constant curvature vs. non-constant curvature targets formed by oriented elements (Baker et al., under review).

General Discussion

In this set of experiments, we compared participants' perception of dot arrays sampled from angular contours with dot arrays sampled from smooth contours. We wanted to better understand how the visual system gets abstract representations of shapes from relatively

impoverished displays of dots. We considered the perceived smoothness of the edges defined by these dots to be one determining factor in their formation. We also wanted to better understand abstract shape representations in general. Using displays in which the contour is not physically present was a useful tool for doing so.

The results from our experiments provide strong evidence for an advantage of smooth contours over angular contours for processing form. In Experiment 1, participants' subjective ratings revealed that dots sampled from smooth shapes were more often and more strongly perceived as shapes than dots sampled from angular shapes. In Experiment 2, we tested subjects' ability to mentally rotate an array of dots, with the expectation that dot arrangements that were more easily encoded as shapes would be easier to recognize across changes in orientation. We found that participants judged that two dot arrays were the same more quickly and more accurately when they were sampled from smooth contours. In Experiment 3, we embedded a target arrangement of dots defining a virtual contour among an equal number of distractor dots. We found that subjects were more able to detect smooth virtual contours than angular virtual contours, likely because the shape representation the dots give rise to is simpler and therefore easier to search (see Baker et al., under review for a similar paradigm).

Greater facility in encoding shapes with fewer corners and more curvilinear segments would not be predicted by many other theories of shape and object perception. Much work in middle and high-level vision emphasizes the importance of junctions and non-accidental properties. Geons in Biederman's (1987) work depend crucially on corners and junctions, for example. Under such a theory, we would expect the visual system to be particularly suited to the detection of corners. Indeed, neurophysiological work points to the importance of angular cusps in V4 (Pasupathy & Connor, 2001). Information theoretical work on contour complexity also

predicts that objects with straight edges will be perceptually simpler (Attneave, 1954; Norman, Phillips & Ross, 2001; Feldman & Singh, 2005). Structural information theory makes the same prediction, positing straight line connections between dots are more economical than curvilinear arcs because arcs are a continuation of both length and angle, thus requiring two bits of information for every one bit of information required for straight line connections (Smits & Vos, 1987, personal communication with Leeuwenberg).

Why, then, are shapes with smooth contours easier to encode than shapes with sharp corners? As Bertamini et al. (2019) point out, there are several reasons to expect angular contours would be more easily processed. Angular contours are comparatively simple to compute, requiring only linear interpolation between salient key points of high curvature (Bertamini & Bates, 2013). There may also be evolutionary advantages to registering the shapes of angular contours quickly to assess danger (Bar & Neta, 2006). On the other hand, the evolutionary environment in which our visual system evolved had much fewer straight edges and sharp angles than the one in which we currently live. Even today, research on scene statistics has found that many of the contours people process in their daily lives are made up of smooth curves (Chow, Jin & Treves, 2002). An analysis of scene statistics can only take us so far, however. The visual system may have evolved to process smooth contours because there were more objects made from smooth contours in our visual environment, but we must still determine what specific visual mechanisms confer this advantage.

One possibility is that the primitives from which the visual system builds abstract shape representations more easily describe a shape with smooth contours. We have hypothesized that shape representations are built up from relatively few smoothly joined segments of constant curvature (Garrigan & Kellman, 2011; Kellman, Garrigan & Erlikhman, 2013; Baker et al.,

under review, Baker & Kellman, in prep.). Under this theory, the presence of a corner would always need to be encoded with two segment primitives, but smoothly changing curvature could be capture by a single segment, provided that the variation in curvature was sufficiently small. Even when part of a smooth contour requires multiple constant curvature primitives, the smoothly joined segments might be perceived as belonging to a single part if they are smooth and monotonic. For example, consider the middle and center contour fragments in Figure 13. Both are made up of two different curvature segments, but the fragment made from two smoothly joined curves looks like a single token, while the fragment made from two straight segments does not.

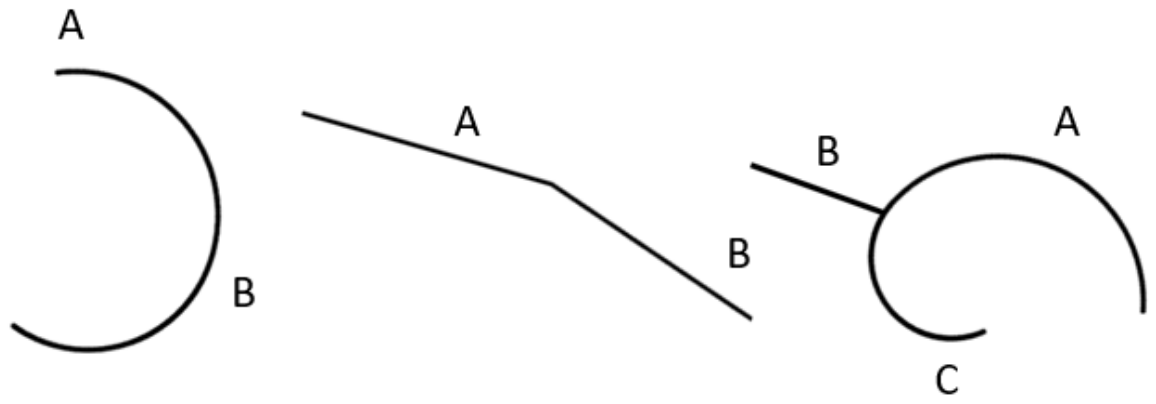

Figure 13. Contours made up of more than one curvature segment. Left: A contour made up of two smoothly joined constant curvature segments. Middle: A contour made up of two straight segments joined at an angle. Right: A contour made up of two smoothly joined constant curvature segments and one straight segment joined at an angle. Individual parts are marked by letters A, B, or C.

The observation that two smoothly joined curvature pieces are perceived as a single part goes back to classic work on good continuation by Wertheimer (1923). He showed a contour similar to the one in Figure 13 (right) and asked people to divide it into two parts, finding that people almost always organized the two smoothly joined curvature pieces together, separating it from the straight segment. In terms of derivatives, we may consider any continuous (unbroken)

contour to have zero order continuity. Wertheimer's examples show that, despite zero-order continuity, a 1st-order discontinuity (undefined first derivative) produces some degree of perceptual segmentation. Figure 13 (right) also illustrates that higher discontinuities, such as the 2nd-order discontinuity where two curves smoothly join (matched slope at the join point), do not produce obvious perceptual segmentation. Evidence from visual search in noise shows that search for a contour segment with a 0-order or first-order discontinuity from other segments is easy, but a segment having 0-order and first-order continuity, but a second-order discontinuity, is effortful, slow, and error-prone (Kellman, Garrigan Kalar & Shipley, 2003). If shapes made up of sharp corners are perceived to have significantly more parts than shapes made up of smoothly connected contours, it follows that they will be more representationally complex and therefore more difficult to encode.

Another reason that shapes with sharp corners may be harder to encode is that the corners are perceived as T-junctions or L-junctions, visual cues that are useful for identifying points at which one object might be occluding another (Ratoosh, 1949; Dinnerstein & Wertheimer, 1957; Rubin, 2001). The presence of a corner might force the visual system to consider the possibility that it is seeing two occluding objects rather than a single object. The additional computations needed to falsify this possibility might slow down form processing.

Conclusion

The results from these experiments suggest that the formation of a shape representation depends on smoothness constraints in the spatial relationships between the dots. These constraints point to a more general phenomenon in shape perception that extraction of curvature is a fundamental process in the formation of an abstract shape representation. Virtual contours that can be described by a relatively constrained set of curvature primitives appear to give rise to

shapes more often, more quickly, and more precisely than virtual contours that are better described by straight segments connected at 1st order discontinuities.

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review*, *61*(3), 183.

Baker, N. & Kellman, P. J. (2020). Constant curvature modeling of abstract shape representation. In preparation.

Baker, N., & Kellman, P. J. (2018). Abstract shape representation in human visual perception. *Journal of Experimental Psychology: General*, *147*(9), 1295.

Baker, N., Garrigan, P., & Kellman, P.J. (2020). Constant Curvature Segments as Building Blocks for 2D Shape Representation. Under review.

Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological science*, *17*(8), 645-648.

Bertamini, M., Helmy, M., & Bates, D. (2013). The visual system prioritizes locations near corners of surfaces (not just locations near a corner). *Attention, Perception, & Psychophysics*, *75*(8), 1748-1760.

Bertamini, M., Palumbo, L., & Redies, C. (2019). An advantage for smooth compared with angular contours in the speed of processing shape. *Journal of Experimental Psychology: Human Perception and Performance*.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review*, *94*(2), 115.

Bouma, H. (196). Perceptual functions. In J. A. Michon, E. G Eijkman, & L. F. de Klerk. *Dutch handbook of psychonomy* (pp. 229-287). Deventer, Netherlands: Van Loghem Slaterus (in Dutch).

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In *Visual information processing* (pp. 75-176). Academic Press.

Chow, C. C., Jin, D. Z., & Treves, A. (2002). Is the world full of circles?. *Journal of vision*, *2*(8), 4-4.

Dinnerstein, D., & Wertheimer, M. (1957). Some determinants of phenomenal overlapping. *The American Journal of Psychology*, *70*(1), 21-37.

Feldman, J. (1996). Regularity vs genericity in the perception of collinearity. *Perception*, *25*(3), 335-342.

Feldman, J. (1997). Curvilinearity, covariance, and regularity in perceptual groups. *Vision Research*, *37*(20), 2835-2848.

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological review*, *112*(1), 243.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision research*, *33*(2), 173-193.

Garrigan, P., & Kellman, P. J. (2011). The role of constant curvature in 2-D contour shape representations. *Perception*, *40*(11), 1290-1308.

Hochberg, J., & Gellman, L. (1977). The effect of landmark features on mental rotation times. *Memory & Cognition*, *5*(1), 23-26.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & cognition*, *13*(4), 289-303.

Kellman, P., Erlikhman, G., & Carrigan, S. (2016). Is there a common mechanism for path integration and illusory contour formation? *Journal of Vision*, 16(12), 311. doi:10.1167/16.12.311

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive psychology*, *23*(2), 141-221.

Kellman, P. J., Garrigan, P. B., Kalar, D., & Shipley, T. F. (2003). Good continuation and relatability: Related but distinct principles. *Journal of Vision*, *3*(9), 120-120.

Kellman, P. J., Garrigan, P., & Erlikhman, G. (2013). Challenges in understanding visual shape perception and representation: Bridging subsymbolic and symbolic coding. In *Shape perception in human and computer vision* (pp. 249-274). Springer, London.

Koffka, K (1931). Psychology of visual perception. In A. Bethe, *Handbook of normal and pathological physiology* (pp. 1215-1271). Berlin: Dessoir (in German).

Koffka, Kurt. "Principles of Gestalt Psychology, 481-493." (1935).

Kovacs, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences*, *90*(16), 7495-7497.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, *6*(4), 225-234.

Lezama, J., Randall, G., Morel, J. M., & von Gioi, R. G. (2016). Good continuation in dot patterns: A quantitative approach based on local symmetry and non-accidentalness. *Vision research*, *126*, 183-191.

Lindemann, E. (1922). Experimentelle Untersuchungen über das Entstehen und Vergehen von Gestalten. *Psychologische Forschung*, *2*(1), 5-60.

Michotte, A. (Ed.). (1964). *Studia psychologica...* Publications universitaires.

Norman, J. F., Phillips, F., & Ross, H. E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception*, *30*(11), 1285-1294.

O'Callaghan, J. F. (1974). Computing the perceptual boundaries of dot patterns. *Computer graphics and image processing*, *3*(2), 141-162.

Papari, G., & Petkov, N. (2005, October). Algorithm that mimics human perceptual grouping of dot patterns. In *International Symposium on Brain, Vision, and Artificial Intelligence* (pp. 497-506). Springer, Berlin, Heidelberg.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *Journal of neurophysiology*, *86*(5), 2505-2519.

Ratoosh, P. (1949). On interposition as a cue for the perception of distance. *Proceedings of the National Academy of Sciences of the United States of America*, *35*(5), 257.

Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception*, *30*(3), 339-366.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701-703.

Shipley, T. F., & Kellman, P. J. (1990). The role of discontinuities in the perception of subjective figures. *Perception & Psychophysics*, *48*(3), 259-270.

Smits, J. T. S., & Vos, P. G. (1986). A model for the perception of curves in dot figures: The role of local salience of "virtual lines". *Biological Cybernetics*, *54*(6), 407-416.

Uttal, W. R. (1973). The effect of deviations from linearity on the detection of dotted line patterns. *Vision Research*, *13*(11), 2155-2163.

Van Assen, M. A., & Vos, P. G. (1999). Evidence for curvilinear interpolation from dot alignment judgements. *Vision Research*, *39*(26), 4378-4392.

Van Den Berg, M. C. J. (2006). * *Grouping by proximity and grouping by good continuation in the perceptual organization of random dot patterns*. University of Virginia.

Van Oeffelen, M. P., & Vos, P. G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recognition, 16*(3), 341-348.

Van Oeffelen, M. P., Smits, J. T., & Vos, P. G. (1985). The perception of a dotted line in noise: a model of good continuation and some experimental results. *Spatial Vision, 1*(2), 163-177.

Wertheimer, M. (1923). Laws of organization in perceptual forms. *A source book of Gestalt Psychology*.

Yin, C., Kellman, P. J., & Shipley, T. F. (1997). Surface completion complements boundary interpolation in the visual integration of partly occluded objects. *Perception*, *26*(11), 1459-1479.

Zucker, S. W., Stevens, K. A., & Sander, P. (1983). The relation between proximity and brightness similarity in dot patterns. *Perception & Psychophysics, 34*(6), 513-522.

Chapter 4: Independent Mechanisms for Processing Local Contour Features and Global Form

Nicholas Baker[1]* & Philip Kellman[1]

1. Department of Psychology, University of California, Los Angeles, Franz Hall 502 Portola Plaza, Los Angeles, CA 90095

*Corresponding author. Please send any inquiries to nbaker9@ucla.edu

Abstract

The human visual system has a robust capability of extracting the global form of an object from a variety of local contour features that often have very little physically in common. We propose a new hypothesis about separate systems for processing high frequency local information along a contour and for encoding global information about the contour's overall shape. We propose that these two systems are independent of each other and process information very differently. While the system encoding information about an object's form represents low frequency contour variations accurately, the local system encodes only a small set of summary statistics to describe typical features of high frequency elements along the contour. In Experiment 1, we tested this hypothesis by sequentially showing pairs of shapes that differed in local features, global features, or both. Support for our hypothesis was found in participants' low sensitivity to changes in local contour features and a lack of additivity for shapes that differed in local features and global features compared to shapes that differed only in global features. In Experiment 2, we tested sensitivity to local and global shape changes once again, this time controlling for the amount of physical dissimilarity between both kinds of shape changes. We found that sensitivity remained higher for global features even with physical similarity equated. In Experiment 3, we compared participants' sensitivity to new sets of contour features with matched statistical properties with new features that differed in frequency and amplitude. Sensitivity was higher when the statistical properties of the contour changed than when new features were generated from the same distribution. We directly tested our hypothesis that local and global properties of a contour are independent features in Experiment 4 using a visual search task. Though local and global shape differences popped out on their own, integrating them together required focal attention. Taken together, these findings support the notion that separate mechanisms process local and global

contour information and that the kinds of information these mechanisms encode are fundamentally different.

What is the identity of the object shown in Figure 1a? The question elicits two distinct but equally automatic responses. On one hand, the object is clearly a cloud, as evidenced by its puffy white texture, its bright reflection of the sun, and its position in the sky. On the other hand, something about the shape of the cloud signals a horse in our perceptual systems. It would be simple to say that surface qualities and context support a cloud percept, while shape cues support a horse percept, but the contour features physically present in the image bear very little resemblance to the set of physical contour features typical in an actual horse. In Figure 1b, we find a horse in a similar pose to the cloud horse and compare the edge map (Canny, 1986) for the same local area in both images. The wisps and curls present in the cloud are visibly absent in a real horse.

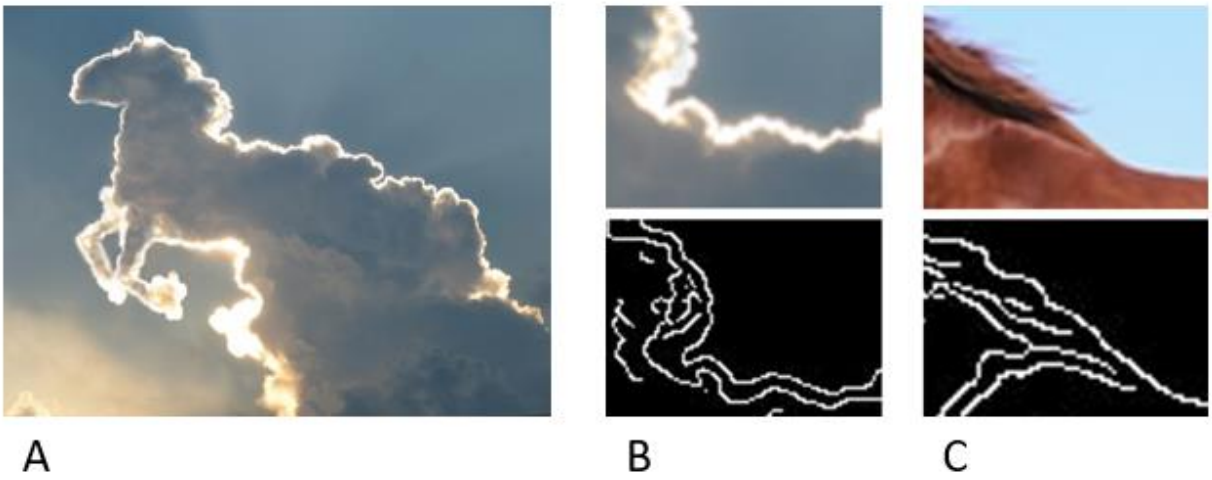

Figure 1. Example of objects with different local contour features but matched global shape. A: Cloud shaped like a horse. **B and C**: Neck and shoulder blades for the cloud horse and a real horse, with corresponding edge maps. The edge maps were generated by MATLAB's implementation of the Canny edge detection algorithm. Cloud image reprinted with permission from https://live.staticflickr.com/5226/5636888777_63246b3359_z.jpgv

Any correspondence between the cloud in Figure 1 and a real horse must be at a higher level of abstraction than the extraction of local contour features. As the Gestalt psychologists observed long ago, the representations we ultimately form of a shape are not a simple

conjunction of the local elements present during sensation (Koffka, 1935). Wertheimer (1923) outlined several principles by which distinct local elements could be organized to form a relational whole. Although similarity was among these cues, the physical characteristics of these individual elements mattered much less than how they were arranged with respect to each other. Under these principles, two contours with very different local elements can be perceived as the same as long as the relations between their constituent elements are sufficiently similar.

There is a great deal of evidence from work on human perception that the visual system extracts global properties of shape before accessing features of individual elements. For example, when viewers are shown an S made up of small H's or vice versa, they first perceive the letter formed by the composition of elements and perceive the identity of the composing elements only later (Navon, 1977). Shape representations also appear to be insensitive to changes in positions of elements provided that the curvature of contour they define is preserved (Baker & Kellman, 2018). Research into the perception and recognition of line drawings has also found that simplified pictorial representations of shape are encoded more rapidly and accurately as the drawing's fidelity to a photographic image deteriorates (Hochberg & Brooks, 1962; Biederman & Ju, 1988). Such findings suggest that the visual system extracts an abstract form from a physical contour, and that when certain abstractions are already present in the image, the computation carried out in perception is simplified.

These findings from human perception stand in stark contrast with some recent studies on object recognition in deep convolutional neural networks (DCNNs). Comparisons between DCNNs' and humans' use of shape in object recognition has shown a large divergence in in each systems' sensitivity to local and global information. Baker, Lu, Erlikhman & Kellman (2020) found that when silhouettes of objects were part-scrambled silhouettes such that many local

contour features were preserved but global shape was destroyed, deep networks continued to classify the altered shapes with equal accuracy and confidence as the unscrambled original images, but when we disrupted local contour features by adding a serrated edge to the boundary of objects while preserving global form, networks' performance deteriorated to chance levels. By contrast, humans had difficulty classifying the part-scrambled images, but no trouble with the images with changed local contour features and preserved global shape (Baker, Lu, Erlikhman & Kellman, 2018). When we introduced a specialized training curriculum to bias networks towards global shape classification, networks learned to filter over larger variations in the contour but developed no sensitivity to an object's global shape (Baker, Lu, Erlikhman & Kellman, 2020). The large influence of local contour features on deep network classification serves as contrasting evidence that humans have special capabilities for separating local and global shape information.

Many plausible models of shape description, however, are not well suited to handle variations in local contour features in a single system. For example, consider theories of part decomposition that separate objects into parts between local concavities (Hoffman & Richards, 1984; Barenholtz, Cohen, Feldman & Singh, 2003). Typical objects like a real horse will be organized into a set of parts that more or less aligns with a semantic part decomposition for legs, neck, body and tail. For the cloud horse, small wisps and bumps around the outline create breaks in good continuation that would give rise to a very different (and more numerous) set of parts than would be seen in any real horse outline. We have proposed a theory of shape perception in which areas of similar curvature are represented by a single segment of constant curvature (Baker, Garrigan & Kellman, 2020; Baker & Kellman, in prep.). If curvatures are analyzed only with fine-scale detectors on the physical contour boundary, a constant curvature segmentation of shapes with high frequency contour information would be made of many tiny primitives and

would certainly not match up with a topographically similar shape in which the high frequency contour features were omitted.

One way the visual system could separate local contour features from global shape is by including a constraint for simplicity when encoding a shape representation. Feldman and Singh (2006) implemented such a method in their Bayesian estimation for shape skeletons. In the classical medial axis transform (Blum, 1973), the skeletal representation of a shape is purely data-driven, and a small bump or protrusion along the contour will always be captured by an axial branch. The maximum *a posteriori* skeletal transform proposed by Feldman and Singh forces the algorithm to tradeoff between simplicity in curvature and number of branches (a prior) and fidelity to the original contour (a likelihood), resulting in a representation that aims to capture the essential topography of a shape without including all local variations to the bounding contour.

Research on shapes formed from radial frequency (RF) patterns have also lent important insight into how the visual system might encode global shape irrespective of variation in local contour features. Radial frequency patterns are sinusoidal modulations of differing amplitude and frequency that can be independently added to a circle to change its shape (see Figure 2) Wilkinson, Wilson and Habak (1998) added RF patterns to circles to evaluate people's sensitivity to global (low RF) and local (high RF) contour features. They found that participants had accurate recognition for shapes with any RF of 6 or less. Recognition for shapes with radial frequencies greater than 6 was considerably more error prone and deteriorated monotonically with larger RFs. Converging evidence for special sensitivity in low RF shapes was later found in a 2IFC task in which subjects discriminated an RF shape from a circle. Sensitivity to a difference from a circle was found to be better than predicted by local probability summation (i.e., the

probability that participants were attending to a region modulated by the RF pattern during presentation) for RFs between 3 and 5, but not for high frequency patterns (Loffler, Wilson & Wilkinson, 2003). The authors theorized that the visual system encodes shape as a combination of radial frequency patterns and that only low frequency RFs are ultimately used in our abstract representation of the shape.

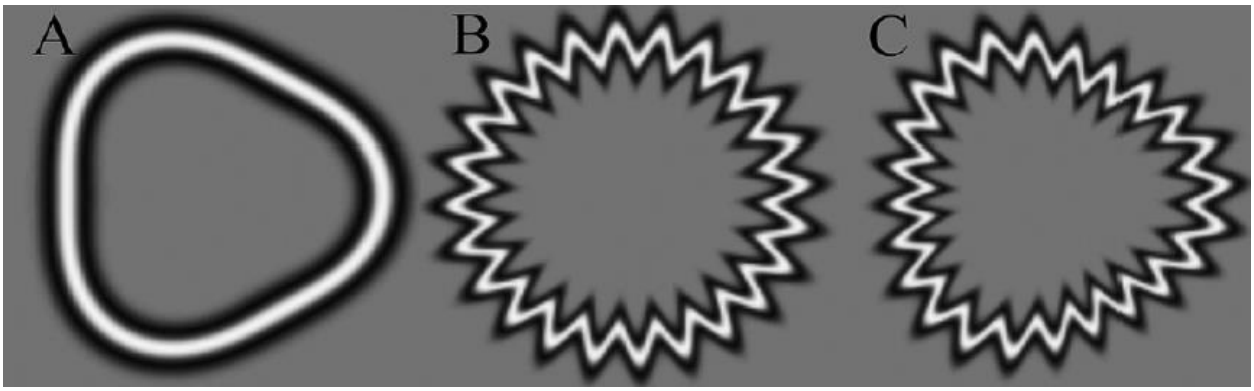

Figure 2. Shapes generated by the addition of radial frequency patterns (from Bell, Badcock, Wilson & Wilkinson, 2007). A: Circle deformed by the addition of a pattern with three cycles (RF3). **B:** Circle deformed by the addition of an RF24 pattern. **C:** Circled deformed by the addition of RF3 and RF24 patterns.

Clearly, the visual system has quite a robust capacity to filter high frequency local contour features out of a representation of shape to support matching across variations that are inconsequential to the object's global topography. It would be a mistake, however, to think that local contour features are thrown out altogether. When we see a cloud that looks like a horse, we do not actually believe we are seeing a horse. This is true even if we remove all surface and context information and attend only to the object's bounding contour, as evidenced by our visual memory for puffs and wisps around the cloud horse's boundary in Figure 1. Any explanation for how we extract global shape from high frequency local contours should also explain what information about local contour variation does get preserved in visual representations.

In the current work, we propose a two-system theory of shape representation that includes separate mechanisms for local and global processing. The local processing system is responsible

for encoding high frequency variations along an object's contour, while the global processing system encodes low frequency topographical information about the shape. In our theory, these systems operate independently from each other such that changes to the high frequency contour features of an object do not affect our representation of its overall shape, nor does our representation of the object's shape interact with our description of local contour features. We posit that the two systems have very different levels of specificity. While the global processing system appears to encode a complete, robust description of the object's overall shape, the local system encodes summary statistics about the local contour features, possibly estimating a distribution from which the contour variations were sampled.

The idea that features are not represented individually but as a group has been proposed in other areas of object representation, such as what people choose to present in line drawings. When drawing a skyscraper, for example, one child chose not to individually draw every window, instead drawing a few accurately and writing "etcetera" for the rest (Arnheim, 1971). Kennedy (1974) found that both children and adults omit repetitive details in line drawings once a few detailed exemplars have been drawn that can be extrapolated to the others. One could well imagine that the information that is represented in visual memory uses similar simplifications to free up perceptual resources for other visual tasks.

Some evidence for independent systems has already been found in research on radial frequency shapes showing that the addition of low frequency features does not affect participants' sensitivity to differences between a high frequency contour and a circle or vice versa (Bell, Badcock, Wilsion & Wilkinson, 2007). Research on processing for different radial frequency patterns has given many valuable insights about differences between local and global shape processing, but the statistics of contours defined in this way tend to be much more

constrained than in typical object contours. For example, the bounding contour of a poodle or a pine tree has much more variety in its local contour variations than would be captured by a small number of RF patterns. For this reason, we developed a different system for generating high frequency contour noise along an object's boundary with fewer statistical regularities than a set of sinusoids.

In Experiment 1, we showed participants two shapes, one after the other, and then had them perform a two-alternative forced-choice task to decide if the shapes were the same or different. We tested for independence between local and global shape processing mechanisms by generating shape pairs that differed in local contour features, in global contour features, or in local and global contour features. We tested whether the inclusion of local and global contour differences conferred an advantage in detecting a difference in shape above what was conferred by one kind of difference. In Experiment 2, we once again used a 2AFC task, this time controlling for the physical similarity between pairs of shapes that differed in local and global contour features. We tested whether people have different sensitivity to local and global contour differences when the physical similarity between the two conditions was matched. The results of Experiments 1 and 2 indicated that local contour features were represented much more coarsely than global features in the visual brain. In Experiment 3, we tested the hypothesis that our descriptions of high frequency shape are statistical rather than fully descriptive. We compared sensitivity to contour changes that were sampled from the same distribution from which the first set of contour features were sampled with sensitivity to changes when the new features were sampled from a different distribution. In Experiment 4, looked for direct evidence for independence between local and global shape processing. We used a visual search paradigm that both provided converging evidence for separation of local and global systems and, following

predictions from Feature Integration Theory, tested a different behavioral prediction made by our independent systems hypothesis. .

Experiment 1

Experiment 1 tested sensitivity to changes in a contour that resulted in different global shape, in different local contour features, or different local and global contour features. We used a forced choice same/different paradigm. We generated novel shape contours, displayed them to participants briefly, and then tested their ability to say whether a second shape was the same or different from the first. Of particular interest to us were differences in sensitivity to global shape changes and to changes to both the local and global features of the shape. We expected that if both kinds of features were processed in a single perceptual system, there should be some additive effect on sensitivity when both features were changed. On the other hand, if they were processed by separate systems and one system dominated the shape recognition task, there might be no added benefit to sensitivity when the other feature was also changed.

Method

Participants

Twenty-four undergraduates (17 female, 7 male, m_{age} = 21.21) from the University of California, Los Angeles participated in Experiment 1 for course credit. All participants had normal or corrected-to-normal vision. The first seven participants completed the study in the laboratory under controlled conditions, while the other 17 completed the study online through Pavlovia due to social distancing orders related to COVID-19. When analyzed separately, similar patterns of results were observed in both the online and in-person groups.

Display and apparatus

The participants we ran in lab were seated 70 cm from a 20-inch View Sonic Graphic Series G225f monitor. The monitor was set to 1024x768 resolution, with a refresh rate of 100 hz. For the online experiment, we instructed subjects to sit a comfortable distance from the screen. Stimulus sizes were dynamically adjusted to cover the same proportion of the screen regardless of participants' display resolution.

Stimuli

All stimuli were shown as black outlines on a gray background. The stimulus was shown in the center of the screen, extending over an average of 37.5% of the horizontal space and 60% of the vertical space on the screen.

Experiment 1 included three conditions, with separate stimuli generated for each condition. In all conditions, a novel shape was generated by moving 12 control points toward or away from the center of a circle a random distance, then fitting cubic splines through the 12 control points in polar space.

In the local change condition, we added contour features to the boundary by moving 80 control points on the shape boundary toward or away from its center. The average distance these control points were moved was $1/10^{th}$ the distance they were moved when generating the global shape. We initially evenly spaced the control points along the boundary, then jittered some of them a small distance so that the spacing was not truly uniform. We then fit cubic splines through the 80 control points to create local contour features as we did for the 12 control points to create global contour features. A schematic for the formation of the stimuli is shown in Figure 3. To create different pairs in the local change condition, we simply inverted the direction of each control point so that bumps that extended away from the center reversed to extend toward

the center and vice versa. This technique preserves as many statistical properties of the features as possible while still introducing a large amount of physical difference to the bounding contour.

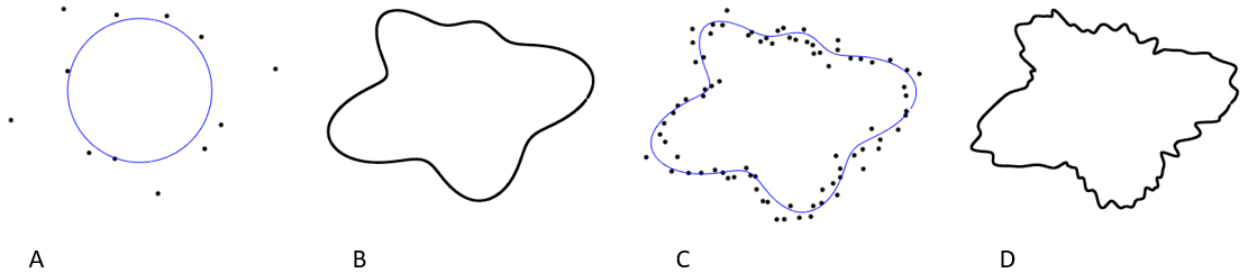

Figure 3. Schematic for global and local shape generation. A: First, a circle is deformed by moving 12 control points away from the center. B: Then, cubic splines are fit through the 12 points. This creates a shape with global features but no local features. C: 80 control points along the shape's contour are moved toward or away from the center of the shape. D: Cubic splines are fit through the 80 new control points.

In the global change condition, we did not add any local contour features to the novel shape. We generated a different pair by moving one of the 12 control points a random distance between 7% and 17% percent of the total length of the contour toward or away from the shape's center. We then randomly selected a point adjacent to the one we moved and shifted it toward or away from the center whatever distance was needed so that the total length of the different pair was the same as the original shape (see Baker & Kellman, 2018 for more detail).

In the combined local and global change condition, we generated pairs of shapes as in the global change condition, but we added local contour features to both members of the pair as in the local change condition. Figure 4 shows pairs of shapes for all three conditions.

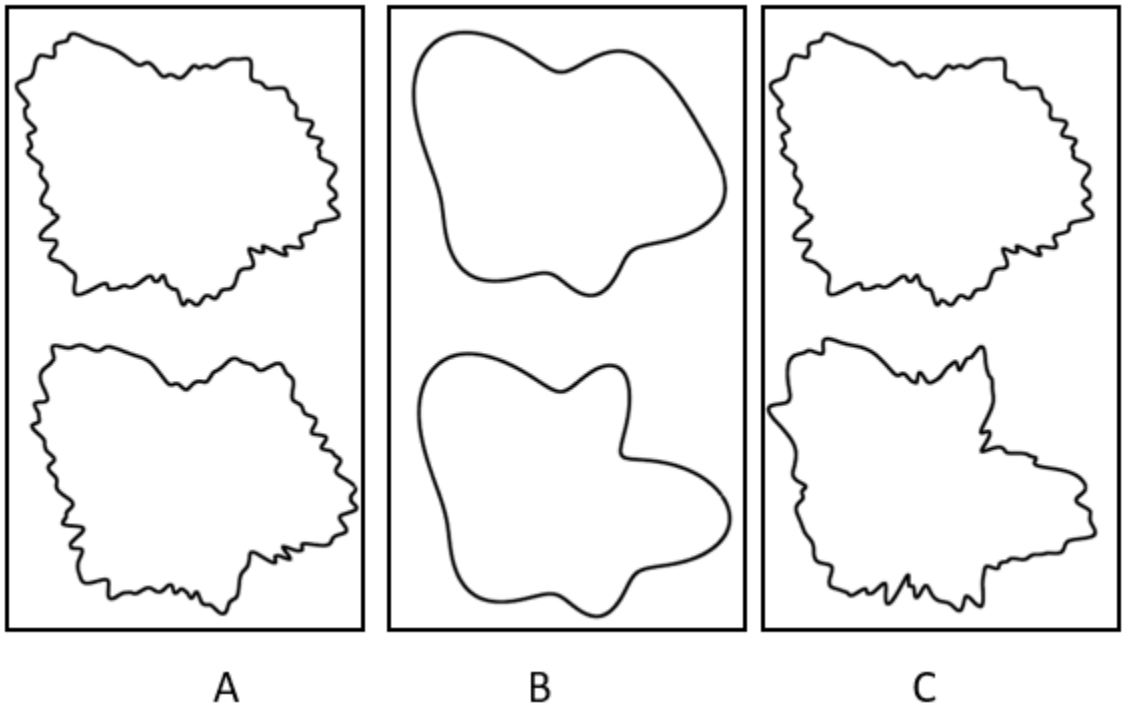

Figure 4. Stimulus pairs for Experiment 1. Left column: Shape pairs that differed in local features. Middle column: Shape pairs that differed in global features. Right column: Shape pairs that differed in local and global features.

Design

The experiment consisted of three conditions, local, global, and local and global, with 80 trials per condition. In half the trials for each condition, the same shape was shown in the first and second presentation. In the other half, a different shape was shown as described above. All trials were randomly interleaved, and there were five practice trials with feedback before participants began the main experiment.

Procedure

On each trial, participants were first shown a fixation cross for 330 milliseconds, after which the first shape was shown for 150 milliseconds. Following presentation of the first shape, a mask was displayed for 500 ms to block any apparent motion cues (Braddick, 1973) or access to a visual icon (Smithson & Mollon, 2006). The second shape was then shown for 1000 ms. The

second shape's orientation was always slightly different from the first, regardless of whether it was a same or different trial. We rotated the shape 10 to 30 in a random direction. After the second shape had been shown, it was masked again and we displayed a response screen in which participants were asked to decide if its outline was exactly the same as the first shape, irrespective of any orientation differences. Participants received no feedback during the main experiment. Sample trials are shown in Figure 5.

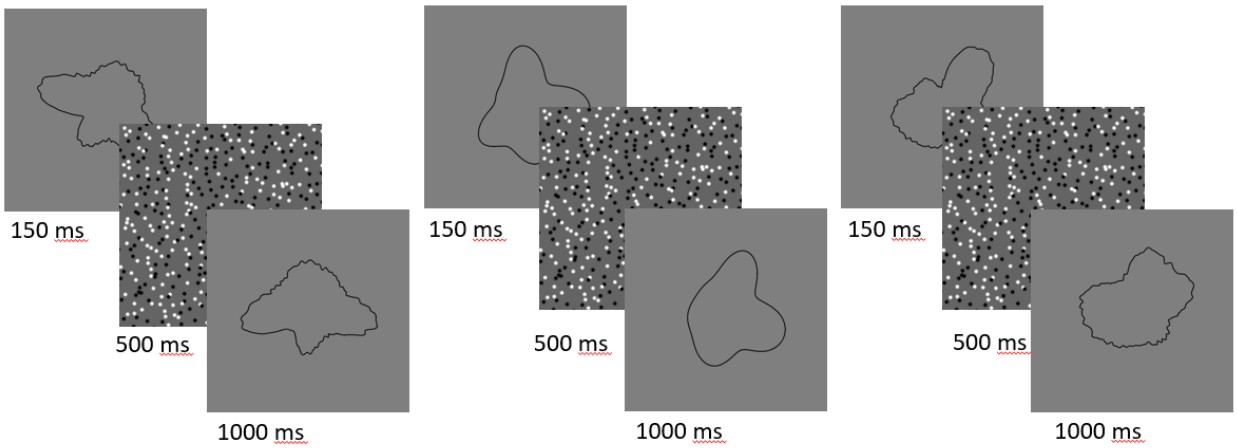

Figure 5. Sample trials for the local, global, and local and global conditions.

Results

We analyzed the results in terms of signal detection sensitivity, where a correct detection that the second outline was different from the first was counted as a hit and an incorrect response that the second outline was different was counted as a false alarm. The same trials were identical for the local and the local and global condition, so we combined them when computing false alarm rates.

Average sensitivity for each of the three conditions is shown in Figure 6. The results show a large difference in sensitivity between changes to the local contour features of the shape and changes to the global topography of the shape. A one way repeated measures ANOVA confirmed a main effect for condition type, $F(2, 46) = 26.32$, $p < .001$, $\eta^2_{partial} = .534$. Paired

samples t-tests confirmed a significant difference in sensitivity between the local condition and the global condition, $t(23) = 9.67$, $p < .001$, *Cohen's d* = 1.97, and between the local condition and the global and local condition, $t(23) = 5.13$, $p < .001$, *Cohen's d* = 1.05. These effects remained significant after correcting for multiple comparisons. A one sample t-test found sensitivity to local shape changes to be greater than zero, $t(23) = 6.83$, $p < .001$, *Cohen's d* = 1.42. A paired samples t-test for the global and the local and global conditions revealed no significant difference, $t(23) = 0.09$, $p = .92$.

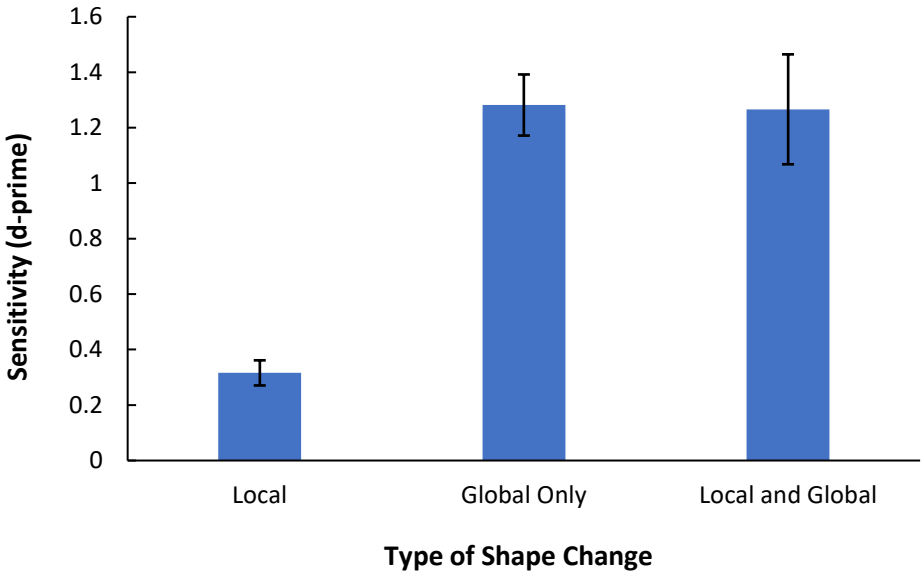

Figure 6. Results from Experiment 1. The x-axis shows the three condition types, and the y-axis shows the calculated sensitivity to each kind of shape change. Error bars show ± one standard error of the mean for each condition.

Discussion

The results of Experiment 1 showed a clear difference in performance for trials in which the shape's global topography changed compared to trials in which the global form remained the same and only local contour features changed. Under our two-system theory of shape representation, local contour features are not represented individually, but as a statistical distribution. The results of Experiment 1 are consistent with such a hypothesis. Note that the in

terms of local features, curvatures, etc., contour features differed at all locations between "different" pairs in the local condition. On the other hand, the statistics of the first and second shape were matched in terms of mean and standard deviation of the amplitude of features as well as the frequency of features along the contour. Sensitivity to actual contour sameness or difference in the local condition was very poor, indicating little encoding of precise contour features. Inspecting the two contours that differed by local features in Figure 4 (leftmost column), it is easy to see that the two are not the same during simultaneous presentation. However, when representing the shape in visual memory, these differences appear to be abstracted over. Performance is consistent with encoding of an ensemble of statistical properties of the distribution from which the features were sampled.

Whereas sensitivity was low for the local change condition, statistical analyses confirmed that participants had significantly higher than zero sensitivity to local feature changes. This forces us to somewhat loosen our hypothesis about local contour features only being described as a few statistical parameters. A more plausible explanation is that people primarily encode statistical information about the local contour features, but they are able to encode a few individual contour features with greater specificity. Non-zero sensitivity in pairs with statistically matched local contour features would then be explained by the probability that participants detected a change in an attended local feature. This notion is somewhat similar to the idea of local probability summation from work with radial frequency patterns (Wilkinson et al., 19998; Loffler et al., 2003; Bell, Wilkinson, Loffler & Badcock, 2009). Our conjecture is that the encoding of individual features is effortful and done only when it would be directly helpful for a perceptual task at hand.

The other major result from Experiment 1 is that participants showed the same sensitivity to changes in only the global features of a shape as they did to changes to both global and local features of the displays. This result is surprising given participants' non-zero sensitivity to local contour features. In a simple summation model, even if the probability of detecting a local change is low, it should still increase participants' overall sensitivity to a change in the local and global condition. For example, if the probability of detecting a local change is 30% and the probability of detecting a global change is 70%, then the probability of detecting either a local or a global change should be 79%. No such additivity was observed in Experiment 1.

One interpretation of this finding is that global information dominates local information when both cues are present. When participants need to detect a change based on only local features, they have some (albeit low) chance of success, however they appear to discard this information when global differences are present. Interestingly, because the conditions were interleaved, participants did not know whether the second shape would have only local contour differences or local and global differences while forming a representation of the first shape. They must, then, have encoded what local information they could about the first shape in both conditions and discarded it only after seeing the second shape. We take this as evidence that the visual system processes local contour features and global shape features independently, only choosing which description to use in its perceptual decision once the second shape has been seen. Possibly, the visual system first checks for different outputs in the global shape system and only compares outputs from the local contour system if no difference is detected.

This interpretation does not fully explain why there is not some performance gain from checking for local differences when the visual system does not detect a global difference in the local and global stimulus condition. The primary answer lies in the suggestion from the data that

the encoding of local contour features does not generally occur; to a first approximation, only some statistical summary is represented. In Experiment 1, where local changes disrupted the local contour position at essentially every location, use of local information for discrimination was still severely limited by the fact that the local variations were sampled from the same statistical distributions. Although performance in the local change condition is consistent with encoding of some specific local contour features, the low d' value associated with local features suggests that, if people responded in an unbiased way, they would report a local contour difference in 56% of trials. Since chance performance is 50%, this means that an accurate detection of a local contour difference only occurred in 6% of trials. There are a number of possible explanations for why observers might occasionally encode some local contour information apart from the statistical summary. We return to this issue in the discussion following Experiment 2 and in the General Discussion. For the moment, the overall conclusion is that performance in Experiment 1 was generally consistent with the global – local divergence of encoding and the idea that the local system apprehends a statistical summary. The results were generally inconsistent with any idea of local encoding that preserves much information about specific contour fluctuations and their locations apart from global shape.

Based on our theoretical framework, we would predict different results if two shapes' local contour features were sampled from *different* distributions. In that case, there should be both improved discrimination for pairs with instances differing in local contour characteristics as well as some evidence of additives effect in a combined local and global change condition. This prediction is tested in Experiment 3 below.

Experiment 2

In Experiment 1, we found that participants were less sensitive to changes in local contour features than they were to global shape. We interpreted this as a difference in the descriptive specificity of independent local and global shape processing systems. An alternative explanation is that the differences between pairs in the global condition were simply larger than differences in the local condition and were therefore easier to detect. In Experiment 2, we test this alternative explanation by equating the physical dissimilarity between local and global shape changes and comparing subjects' sensitivity to each. We predicted that if the two systems were distinct and represented information differently, then subjects' sensitivity to global shape changes should still be higher than their sensitivity to local contour feature changes, even if the physical dissimilarity was the same.

Method

Participants

Eighteen undergraduates from the University of California, Los Angeles (12 female, 5 male, m_{age} = 21.5) participated in Experiment 2 for course credit. Sixteen of the participants in Experiment 2 also participated in Experiment 1. All participants had normal or corrected-to-normal vision and completed the study online through Pavlovia.

Display and Apparatus

Since the experiment was conducted online, display conditions varied slightly from participant to participant. Subjects were instructed to sit a comfortable distance from the screen and stimulus sizes were adjusted to cover the same proportion of the screen regardless of participants' display resolution. We allowed these variations for obvious practical reasons during the Covid-19 pandemic, but also because the perceptual abilities under study here should be robust across a range of ordinary screen sizes and viewing distances.

Stimuli

All stimuli were shown as black outlines on a gray background. The stimulus was shown in the center of the screen, extending over about 37.5% of the horizontal space and 60% of the vertical space on the screen.

Pairs of locally and globally different shapes were generated as in Experiment 1. The only difference in how shapes were generated was that we reduced the amount the control point was shifted in the global condition from 7-17% in Experiment 1 to 2.7-6.7% in Experiment 2. This was done to better equate the physical contour difference between local and global shape pairs. To that end, we also used a measure of physical contour similarity to ensure that the shape pairs with local contour differences were as dissimilar from each other as the shape pairs with global contour differences. Similarity was measured as the ratio of the overlapping areas to the non-overlapping areas for both contours. The equation we used was as follows:

$$\frac{\frac{\text{Shape 1 and 2 overlap}}{\text{Total area of Shape 1}}+\frac{\text{Shape 1 and 2 overalap}}{\text{Total area of Shape 2}}}{2}$$ (see Chapter 1 (Baker & Kellman, in prep.) for more detail).

We generated 120 shape pairs for each of the two conditions. The average total shape difference was 4.52% for pairs of shapes in the global change condition and 4.62% for pairs in the local change condition. Sample shape pairs for Experiment 2 are shown in Figure 7.

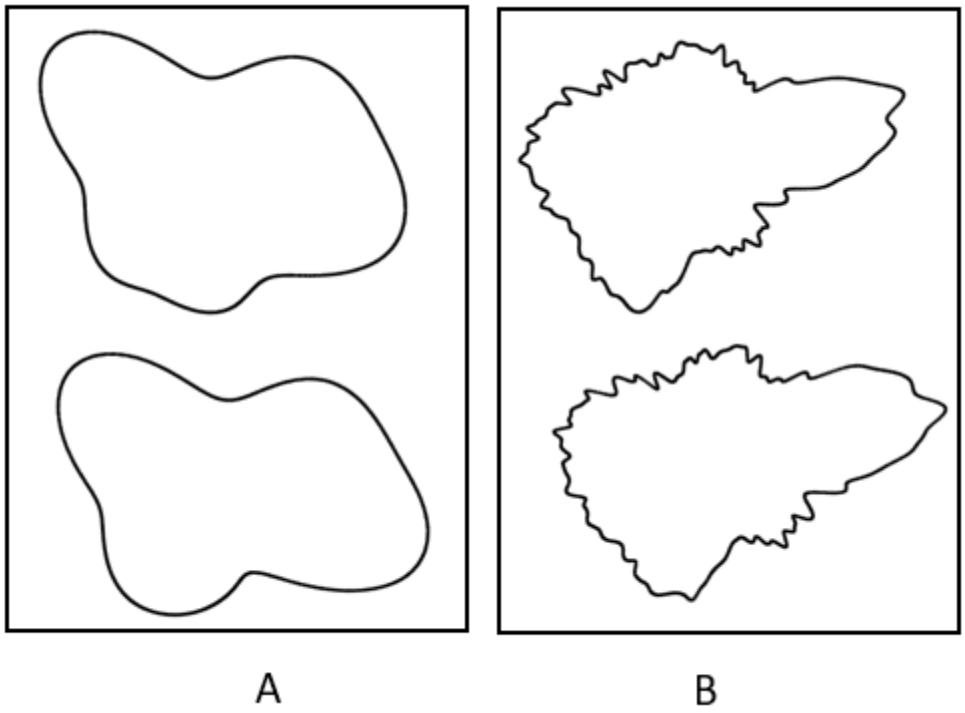

Figure 7. Shape pairs for Experiment 2. Left column: Pairs of shapes that differed in global features (A). Right column: Pairs of shapes that differed in local features (B).

Design

Experiment 2 had two conditions: shapes that differed in local features and shapes that differed in global features. There were 120 trials in each condition. Local and global trials were randomly interleaved. There were five practice trials with feedback before the main experiment began.

Procedure

The procedure for Experiment 2 was the same as in Experiment 1 except that there were no conditions in which both local and global contour features changed. If local contour features were varied the instances in a stimulus pair, then the second stimulus always had the same overall shape and could differ only in local contour features. Likewise, if the first stimulus had no local contour features, the second stimulus could differ only in global shape properties.

Results

As in Experiment 1, the results were analyzed in terms of sensitivity, where a hit was a correct detection of a shape change and a false alarm was an incorrect report of a change in shape. Sensitivity to both conditions is shown in Figure 8. A paired samples t-test confirmed a significant difference in sensitivity to global shape differences vs. local shape differences, $t(17) = 7.15$, $p < .001$, *Cohen's D* = 1.69.

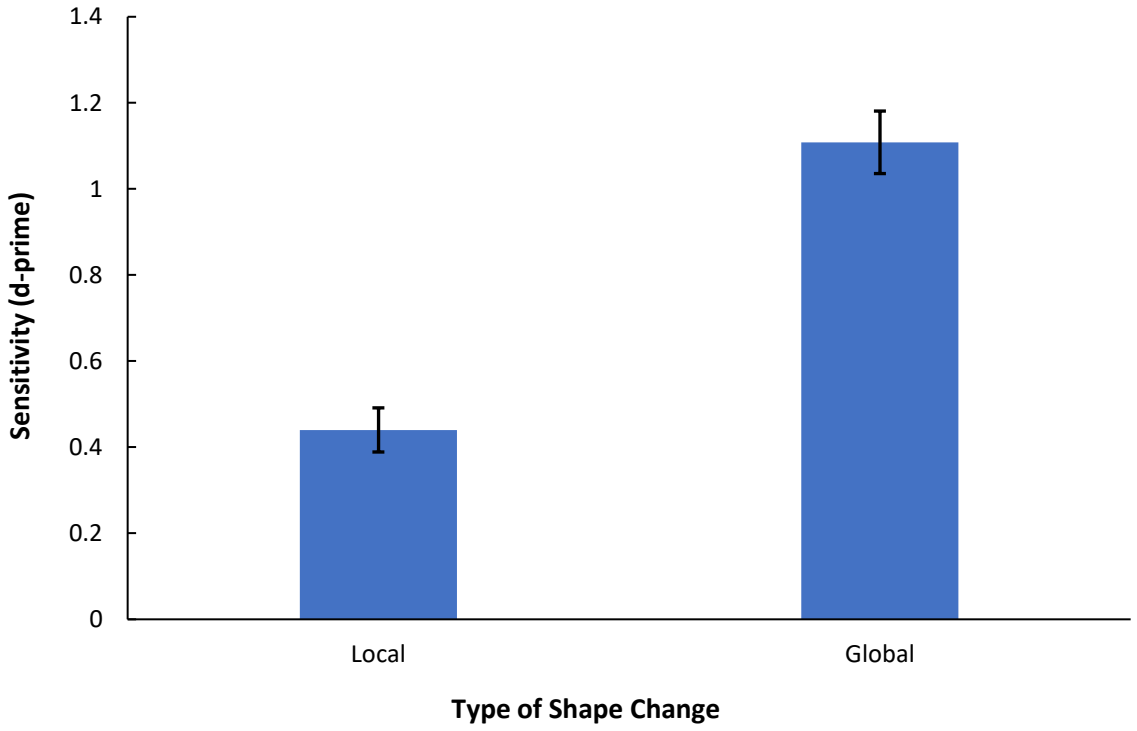

Figure 8. Results from Experiment 2. Error bars show the standard error for each condition.

Discussion

In Experiment 2, we once again compared sensitivity to local and global shape changes. We created pairs of shapes that differed in either local contour features or global form and equated the physical similarity between pairs in both conditions. These were tested in a sequential forced-choice same/different paradigm. Despite having equally different physical

contours, pairs in the local change condition were significantly less discriminable from each other than pairs in the global change condition. These results clarify our interpretation of the results in Experiment 1. Participants' lower sensitivity to local shape changes cannot be explained by a smaller amount of overall physical contour difference, but by a perceptual difference in how local features and global features are encoded. We would expect that, if one system processed both local and global contour information in the same way, then sensitivity to a difference in shape should depend only on the physical similarity of the two shapes. On the other hand, if local and global contour features are processed by distinct systems, the kind of information encoded in one system might be more discriminative between shape pairs than the other.

Since, as in Experiment 1, local contour features were changed by inverting the curvature polarity of the bumps, the statistical properties of locally different shape pairs should have been very similar. Participants' low sensitivity to local contour feature changes is consistent with the visual system encoding a statistical description of local contour features, but not encoding individual features specifically. It is true that sensitivity in local trials was significantly different from chance, but d' values of 0.4 suggest that participants only detected the change in 8% of locally different trials. Possibly, some task-specific strategy or serendipitous alignment of local and global features allowed participants to answer accurately in the small proportion of trials in which a local change was detected. In both Experiments 1 and 2, a local contour change involved a change in all the contour's all local features. If participants were encoding specific local information in any but a very small percentage of trials, their sensitivity to local changes should have been much higher, as specific comparison of any set of local contour features would lead a detection of a contour difference.

Experiment 3

Experiments 1 and 2 supported our global – local theory by furnishing evidence that local contour feature changes were poorly detected. We interpreted this detection difficulty as being due to the use of the same summary statistics for the local contour features for non-identical instances in the local trials. Our interpretations thus far are also consistent with other possible interpretations. One is that there is a single general contour shape processing system, but that our local perturbations were relatively small changes, whereas our "global" differences were larger. Exp. 2 equated the total amount of change for global and local conditions, but there could be other metrics on which these differed, such as the largest local region of change. In general, the failure to distinguish two displays sharing the same contour statistics but having different local features is consistent with our statistical view of local encoding, but it is also consistent with rather poor encoding in general of local contour features.

In Experiment 3, we addressed this issue directly by investigating participants' sensitivity to changes in local contour features in the presence of changes to their summary statistics. Our hypothesis was that when the new set of contour features was sampled from a different distribution, participants would be more sensitive to the change than when the new set of features was sampled from the same distribution.

Method

Participants

Twenty-one (18 female, 2 male, 1 gender not specified, $m_{age} = 20.9$) participants from the University of California, Los Angeles completed this study online for course credit. All participants had normal or corrected-to-normal vision.

Display and Apparatus

Since the experiment was completed online, screen size and distance were not fixed. The experiment adjusted the absolute size of stimuli so that it covered the same proportion of the screen for different monitors.

Stimuli

All stimuli were shown as black outlines on a gray background. The stimulus was shown in the center of the screen, extending over 37.5% of the horizontal space and 60% of the vertical space on the screen.

We generated four kinds of shape pairs for comparison in Experiment 3: matched statistics, different frequency, different phase, and different amplitude. In the matched statistics condition, we sampled a new set of contour features from the same distribution from which the features for the first member of the pair was sampled. Both sets of contour features had 80 control points and the mean of the distribution from which amplitudes were sampled was the same. We also wanted to ensure that shape pairs were not merely sampled from the same distribution but did in fact have matched statistics themselves. To that end, we computed the mean and standard deviation of the amplitude of bumps for both shape pairs and resampled until the differences in their means and standard deviations were both less than 0.01.

In the different *frequency* condition, we manipulated the number of bumps along the contour by a Weber fraction of 1.5, such that the original stimulus with 80 bumps was paired with another shape that had either 53 or 120 bumps. The mean and standard deviation of the amplitudes were matched as in the matched statistics condition.

In the different *amplitude* condition, the amplitude of the bumps were increased or decreased by a Weber fraction of 1.5. The frequency of the bumps was kept the same as the first member of the pair. Sample shape pairs for all four conditions are shown in Figure 9.

In the different *phase* condition, we shifted all the local contour features along the contour 1.67% of the contour's total length. This had the effect of preserving all frequency and amplitude statistics but changing the spatial relationships between the local and global features. The local features were also moved so that their peaks and troughs were approximately halfway between peaks and troughs in the first member of the pair. Whereas the prediction from a summary statistics view of local contour processing is that changes in the amplitude and frequency would change the contour statistics sufficiently to enhance discrimination, the prediction for phase was the opposite. The essence of the local contour statistics hypothesis is that the visual system does not encode the local contour orientations or fluctuations at locations generally. If so, then preserving frequency and amplitude statistics, but moving the positions of particular features to different places along the contour (phase shift) should *reduce* discriminability.

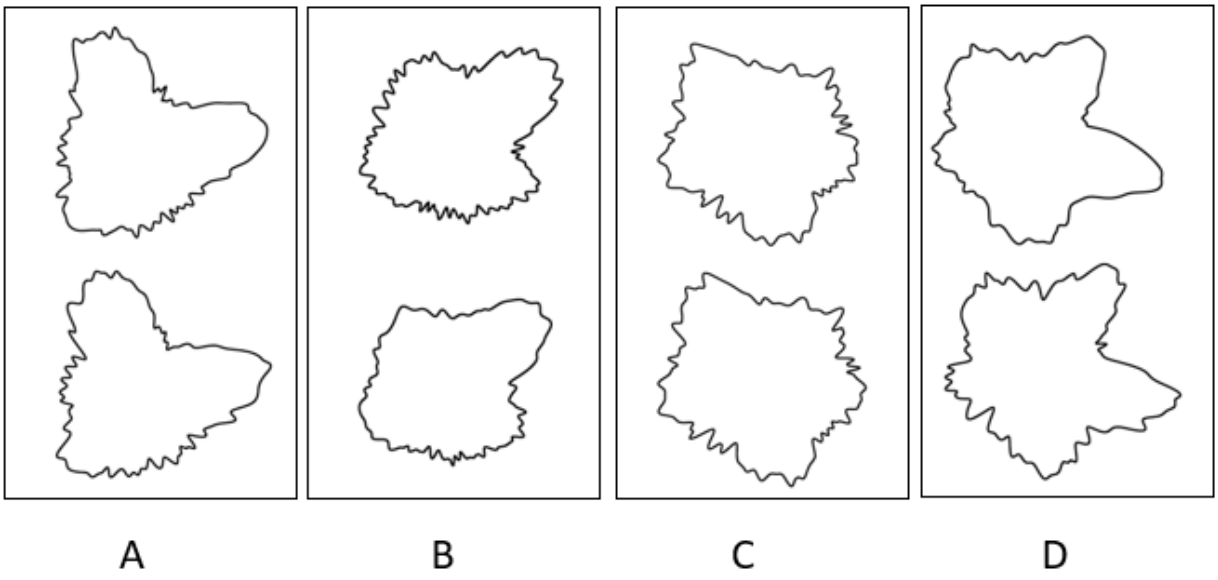

Figure 9. Sample shape pairs for Experiment 3. Columns from left to right: Matched statistics (A), different frequency (B), different phase (C), different amplitude (D).

Design

The experiment consisted of four change conditions with 40 trials per condition plus another 160 no-change trials. All conditions were randomly interleaved. Participants completed eight practice trials before beginning the main experiment. At least one trial from each of the four conditions was completed during practice.

Procedure

Apart from the difference in stimuli, the procedure and instructions for Experiment 3 were identical to those used in Experiments 1 and 2. Participants were shown two shapes, one after the other, then asked to determine whether the second shape was exactly the same as the first shape apart from a difference in orientation.

Results

As in Experiments 1 and 2, we analyzed results in terms of sensitivity, where a hit was a correct detection of a change in shape and a false alarm was an incorrect report of a change in shape. Since there were no differences in the *same* trials between conditions, false alarm rates used to calculate d' were the same for all four conditions.

The primary results are shown in Figure 10. A one way repeated measures ANOVA confirmed a difference in sensitivity between conditions, $F(3,60) = 28.84$, $p < .001$, $\eta^2_{partial} = .59$. Paired samples t-tests revealed a significant difference between the amplitude change and frequency change conditions, $t(20) = 2.96$, $p = .008$, *Cohen's D* = 0.67, a significant difference between frequency change and matched statistics, $t(20) = 2.47$, $p = .022$, *Cohen's D* = 0.52, a significant difference between amplitude change and matched statistics, $t(20) = 5.35$, $p < .001$, *Cohen's D* = 1.16, and a significant difference between matched statistics and phase change, $t(20) = 4.33$, $p < .001$, *Cohen's D* = 0.97. After correcting for multiple comparisons, the difference between frequency and matched statistics became marginally significant.

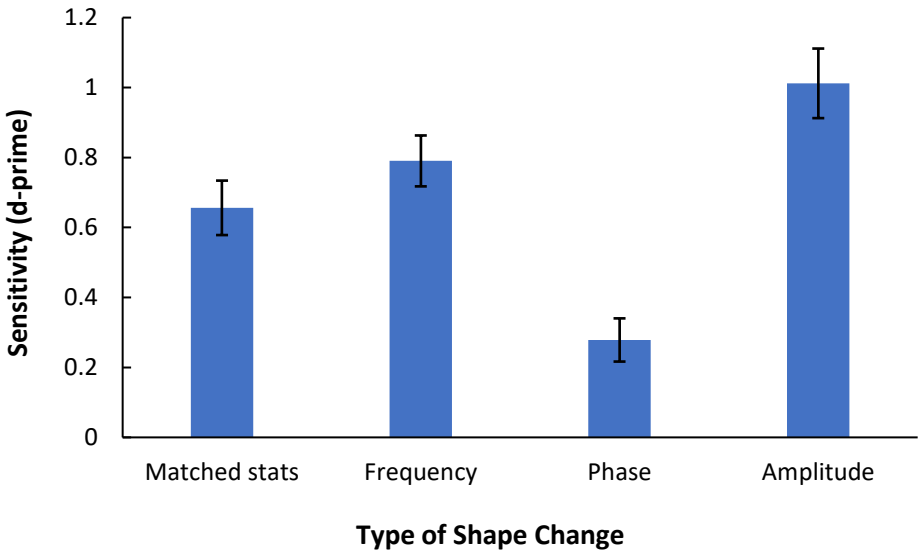

Figure 10. Sensitivity Results from Experiment 3 by Stimulus Condition. Error bars show ± 1 standard error of the mean for each condition.

Discussion

Experiment 3 tested the hypothesis that the system responsible for processing local contour features primarily encodes information about the distribution from which features are sampled, not the properties of individual features. We hypothesized that the statistical information people encode about high frequency contour features might include information related to the mean and standard deviation of the amplitude of features and the frequency of features along the contour. We therefore predicted that participants would have the greatest sensitivity to local contour feature changes that included a change in the frequency or amplitude of the features, and that participants should be relatively insensitive to changes that did not affect the summary statistics of the features, such as resampling from the same distribution or shifting the same features a small amount along the contour.

Our results broadly aligned with these predictions. Sensitivity was highest for amplitude and frequency differences and lowest for phase differences and matched statistics, more closely reflecting participants' sensitivity to a global shape change in Experiments 1 and 2. This suggests

that outputs from the processing system responsible for local features is much more sensitive to differences in the statistical distributions from which local features are sampled than specific descriptions of individual features. These results show that local contour perturbations on the order of those used in Experiments 1 and 2 are not simply poorly encoded; rather, discrimination is poor when local contour features are changed but summary statistics are preserved. In Experiment 3, when summary statistics were varied, performance was reliably better than in the earlier experiments.

Participants performed better for changes in the mean amplitude of contour features than changes in the frequency of features, and both of these were better detected than local feature differences that had matched statistics. In our design, we equated the Weber fraction for differences in amplitude and frequency, using a ratio of 1.5:1 for both conditions. It appears that the local processing system's sensitivity to the frequency of features is coarser than its sensitivity to the amplitude of local features, resulting in less reliable detection of contour changes brought about by a change in frequency.

Sensitivity for contour features that were different but statistically matched to the first shape was 0.66. Performance was better for this condition than was observed in Experiment 1 (d' = 0.32) or Experiment 2 (d' = 0.44). We sampled an entirely new set of contour features for each stimulus pair with locally different instances in Experiment 3, whereas in the previous experiments we inverted the polarity of all contour features. We did this to test a more specific hypothesis about what statistical properties of the local contour the visual system represents. Statistics were matched in terms of mean amplitude, variance of the amplitude, and frequency. These statistics appear to play a large role in subjects' sensitivity to a contour change, but people's higher sensitivity to contour differences in Experiment 3 could suggest that the system

processing local contours encodes other kinds of statistical information about the local features in addition to those we matched. For example, the visual system might encode more information about the distribution of curvatures within a local contour feature.

Despite the use of entirely new contour features for each member of a "different" pair, pairs with different frequency and amplitude statistics were better discriminated. Conversely, and particularly striking, is how *low* participants' sensitivity was to pairs in which the local features were phase-shifted a small distance along the shape's contour. Unlike the other three conditions, the features in the phase condition were otherwise identical to those in the first display, so any statistical differences we may not have accounted for in the matched statistics condition would be precisely matched in phase-shifted shape pairs. At the same time, if local contour features were represented precisely in the same system as global features, we would expect very high sensitivity to changes in phase as the shift along the contour alters relationships between local and global features. Performance in the phase change condition suggests independence between local and global contour features, as there appears to be no precise binding between local features and their position on the shape's global structure.

Experiment 4

In Experiments 1-3, we found evidence that the visual system is more sensitive to the global shapes of objects than to high frequency contour features along an object's boundary. While global shape representations appear to include descriptions of global features and relations of parts, local contour features are seemingly described by a small set of statistical properties. Differences in processing for local and global shape information suggests that they are handled

by distinct systems. In Experiment 4, we developed a convergent measure and additional direct test for the independence of local and global shape features by measuring visual search time.

In visual search, targets that differ from distractors by a single feature tend to "pop out" from the search array, resulting in a search time that is independent of the number of elements in the array. However, when the visual system must integrate two independent features, such as shape and color, search time becomes serial, increasing with the size of the array (Treisman Treisman & Gelade, 1980; Quinlan & Humphreys, 1987; McElree & Carrasco, 1999). According to Treisman and Gelade's (1980) Feature Integration Theory, basic visual features are extracted automatically and in parallel across the visual field, but independent features of an object can only be integrated together with focused attention.

Experiment 4 tested whether local and global aspects of shape are processed together or independently in human perception. If descriptions of local contour features and global form are represented in independent systems, we reasoned that models of feature integration from visual search might predict that they should require focused attention to be integrated together. If, on the other hand, our theory is incorrect and local and global features are processed in the same system, search time for a conjunction of local and global features should take no longer than search time for one feature or the other provided total difference is equated.

Method

Participants

Twenty-one (8 female, 11 male, m_{age} = 31.9) people participated in this study. About half of the participants completed the study for course credit. The other half volunteered to complete the study without compensation. All but two participants were naïve to the purpose of the

experiment before completing it. No significant differences were found in the data with these two participants omitted. All participants had normal or corrected-to-normal vision.

Display and Apparatus

All participants completed the experiment online and had variable screen sizes. They were instructed to sit 1.5 times the diagonal length of their screen away from the monitor while completing the experiment but were also told they could adjust if this distance was uncomfortable for them.

Stimuli

All shapes were shown as black outlines on a white background. Each shape extended over 16.7% of the horizontal space on the screen and about 22% of the vertical space. In each trial, there was a target shape and four, eight, or twelve shapes in a search array. The target was always shown in the top left of the screen outlined by a blue square. In target-present conditions, there was also an identical target in the search array. In the target-absent conditions, no shape in the search array was identical to the target. The shapes in the search array were displayed in a grid. The four-element search array was in a 2x2 grid, the eight-element search array was in a 4x2 grid, and the twelve-element search array was in a 4x3 grid. Shapes in each search array were positioned so that the mean distance from participants' fixation was equated for all three array sizes. All elements in the search array were rotated by the same magnitude and direction 10-30 degrees off from the target exemplar in the top left.

We had three different search array conditions: local, global, and conjunction. In the local condition, all shapes in the search array had the same global form as the target, but only the target had the same local contour features. The distractors all had different local contour features from each other, but the frequency of their contour features was matched, while the target had a

different frequency of features. In half the target-present trials, the target shape had 25 bumps along its contour while the distractors had 80 bumps, and vice versa in the other half. In terms of actual contour features, all shapes in the array were unique, but in terms of statistics of features, only the target was unique. This is different from most visual search studies in which distractors are typically uniform or fall into a small number of categories with identical items in each. Our design leverages the idea that pop-out might still actually be possible based on the common contour statistics of distractors, despite the actual uniqueness of the local contours of every item in the display (for displays where local information was varied between target and distractors).

In the global condition, both the target and all shapes in the search array had the same set of local contour features added to them. The local contour features had 25 bumps in half of the trials and 80 in the other half. The target always had a different global form than any of the distractor. The distractors all shared the same global form and were identical to each other.

In the conjunction condition, the target had a unique combination of local contour features and global form. All shapes in the array had added local contour features, half of which had the same frequency as the target and half of which had a different frequency. Likewise, half the shapes in the array had the same global form as the target and half had a different global form. No distractor had both the same global form and the same frequency of local contours as the target. A sample search array for each condition is shown in Figure 11.

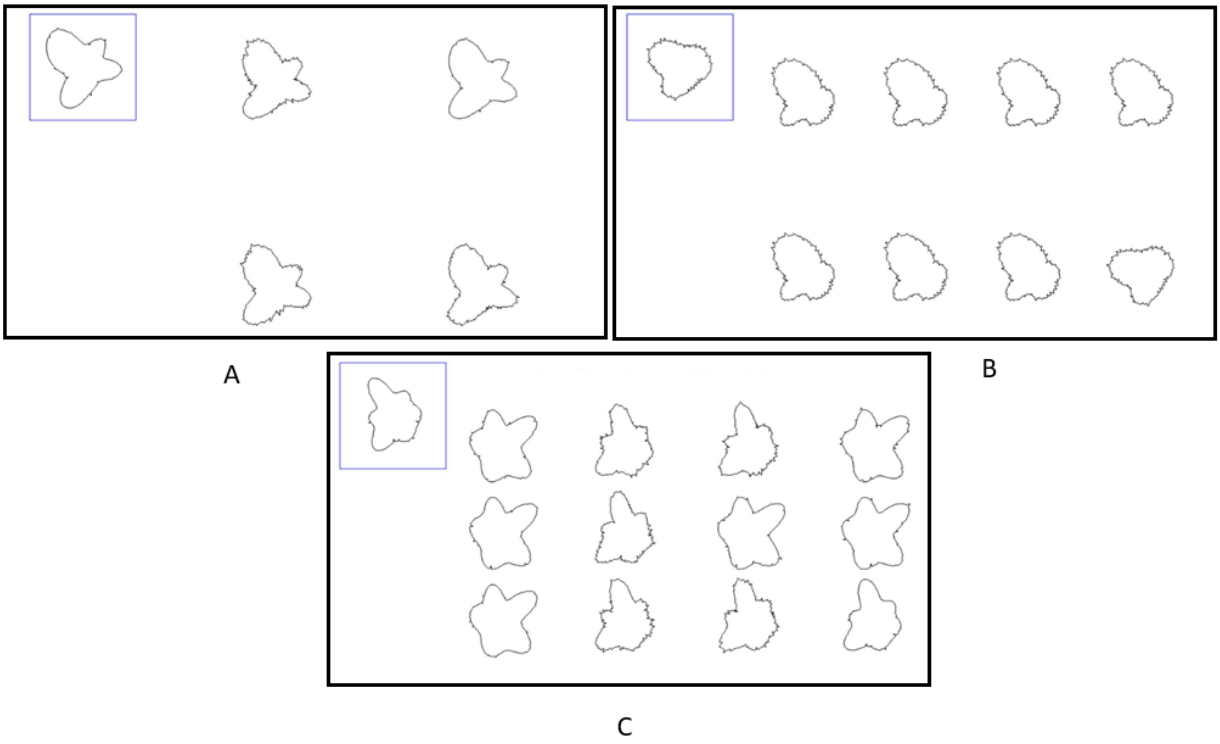

Figure 11. Sample trials for Experiment 4. A: A four-element search array in which the target has different local contour features than the distractors. B: An eight-element search array in which the target has different global form than the distractors. C: A twelve-element search array in which the target has a unique combination of local features and global form. The black rectangles are added to separate trials. The blue rectangle within each condition was shown during the trial to show subjects the target they were supposed to look for in the search array.

Design

Experiment 4 had three by three (condition type x search array size) design. Different condition types were done in blocks with randomly interleaved search array sizes in each block. Each block consisted of 20 target-present trials plus 20 target-absent trials for each of the three array sizes. The order of the three blocks was randomized for each subject to eliminate any systematic fatigue or practice effects. Participants completed six practice trials before beginning the main experiment.

Procedure

On each trial, participants were first shown the search target shape by itself in the top left of the screen enclosed by a blue square. The target remained on the screen for 1500 ms. It then

disappeared and a red fixation cross was presented in the center of search array for 500 ms. The fixation cross disappeared, and the search array was shown on the screen. The target in the top left was also shown with the search array for participants to use as a reference if needed. Participants were instructed to decide if the target was present or absent from the search array. They were told to respond affirmatively as soon as they had found the target, but to check all shapes in the array before reporting that the target was absent. Participants were given feedback after each trial telling them if the target was present or absent.

During practice, we showed each of the three different target conditions and each of the three different array sizes. Feedback in practice was more detailed than during the main experiment. After participants responded, the target (if present) was circled in red and an explanation about how the target differed from the distractors was given. This explanation was different for each of the three conditions.

Before each block in the main experiment, participants were given a brief explanation of how the target would differ from distractors in the next block of trials. We did this to reduce any confusion participants might have in what to look for during the first few trials in the new condition.

Results

Two subjects were removed because their accuracy was below 65% in all three conditions. For the remaining subjects, we analyzed time to respond as a function of array size for the local, global, and conjunction condition. The results are shown in Figure 12. There appears to be a clear linear function showing response time increasing with set size in the conjunction condition. There were much flatter search functions across different array sizes in the local and global conditions. Following Quinlan and Humphreys (1987) we performed a

repeated measures ANOVA and tested for a linear component in within-subjects contrasts. We found a significant linear component for the local condition, $F(1,18) = 16.39$, $p = .001$, and for the global condition, $F(1,18) = 13.92$, $p = .002$. Although the linear component was significant, the variance explained by the linear component for local and global conditions was small, 40.6% and 43.6% respectively. The slope for the two conditions was quite shallow. Search time increased by 46 ms per item in the local condition and 53 ms per item in the global condition. These increases are minimal relative to the intercept terms for each condition (2047 ms and 1869 ms for the local and global condition respectively). When the target had a unique conjunction of local and global features, ANOVA confirmed a much more substantial linear relationship between response time and set size, $F(1,18) = 40.51$, $p < .0001$. The linear relationship explained a much higher proportion of the variance in the conjunction condition (72.2%), and the ratio of the slope (258 ms per item) to the intercept was more than five times larger than in either the local or global condition (.15 vs. .022 and .028, respectively).

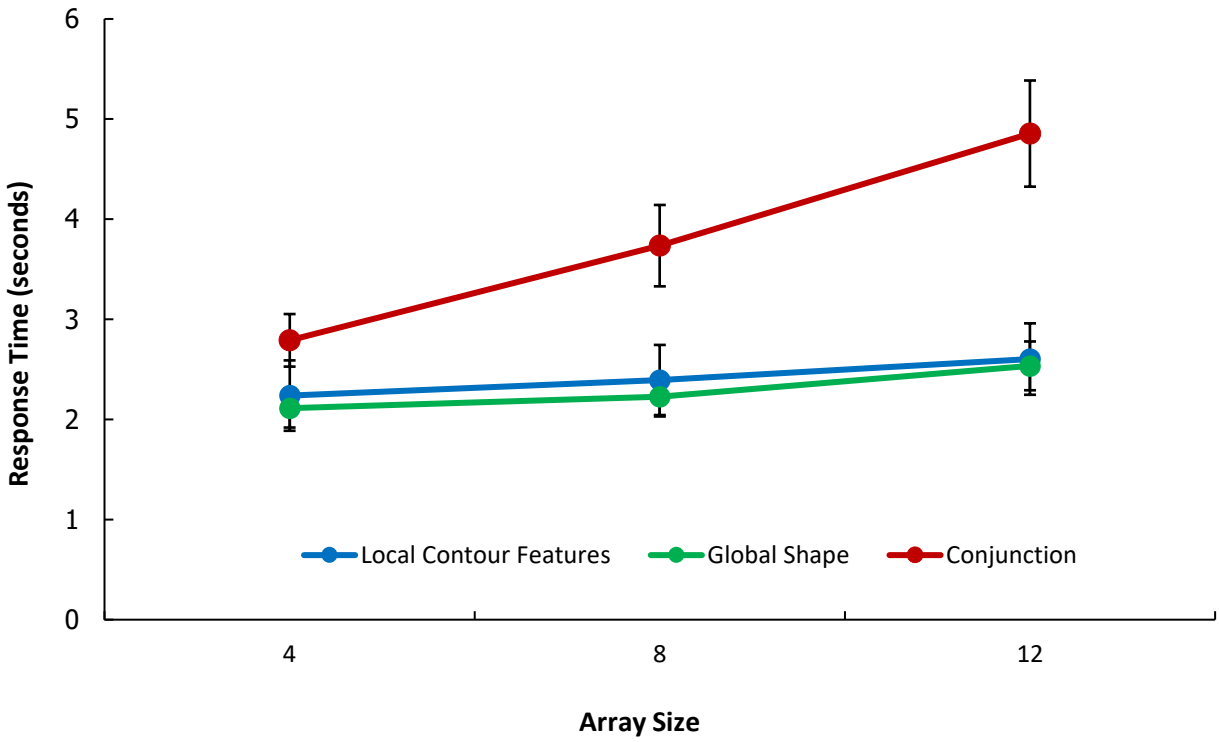

Figure 12. Response times as a function of array size in Experiment 4. Error bars represent the standard error for each condition.

Differences in response time cannot be accounted for by a speed accuracy tradeoff. In our data, accuracy for the local, global, and conjunction condition was 93.7%, 95.4%, and 95.1% respectively. A repeated measures ANOVA confirmed that there were no significant accuracy differences between the conditions, $F(2,36) = 0.61$, $p = .512$. Some research has found a difference in accuracy for conjunction search trials compared to feature search trials (Treisman, 1993; Wolfe, 1994), but we found no evidence for a loss in accuracy when searching for a conjunction of local and global shape features compared to searching for a single feature.

Discussion

Experiment 4 compared visual search time for shapes that differed in local contour features and/or features of global form. Our hypothesis was that local and global features are

distinct and processed independently, and focal attention should be required to integrate both features together (Treisman & Gelade, 1980). An alternative hypothesis is that local and global features are processed by the same system and that global features are simply larger local features. Under this hypothesis, we might expect search time for a conjunction of local and global features should not be different from search time for local or global features on their own. Participants had flat or nearly flat search times for both local contour features and global form. This suggests that they are single features that can be searched for in parallel across the visual field. Although search times did increase slightly with larger array sizes, the increases were small enough relative to the intercepts that they are most likely explainable by the presence of more retinally eccentric shapes in the larger array (Palmer, 1995; Scialfa & Joffe, 1998; Eckstein, 2011) and/or crowding effects from shapes being more close together in larger arrays (Vlaskamp & Hooge, 2006).

The speed at which participants could detect the target in the local difference condition is especially remarkable because all shapes in the array, including distractors, had different sets of local contour features. What made the target shape pop out among the distractors was not the uniqueness of its high frequency features but a unique set of statistical properties of the high frequency features. Shapes that had matched statistical properties, but different local elements, could only be differentiated by close scrutiny, but shapes with different statistical properties popped out in the array. In Experiment 3, we found a (statistically adjusted) marginal difference in discriminability for shape pairs with matched statistics and shape pairs with a frequency ratio of 1.5:1. We conjectured that the visual system's sensitivity to differences in frequency might depend on a larger Weber fraction. Experiment 4 supports that conjecture—statistical differences in frequency result in a pop out when the ratio is increased to 3.2:1.

Support for our hypothesis that local and global shape features are distinct and independent comes from the large increase in search time as a function of array size observed in the conjunction condition. Pop out ceases when detection of the combined features is required in the search task. Even if one argued that the single-feature visual search tasks in the local and global condition are not truly parallel, the difference in slope for the conjunction condition indicates an integration of two independent features. Under the alternative hypothesis that local and global features are processed in the same system, we would expect the degree to which visual search varies with array size to depend on the similarity between the target and the distractors (Pashler, 1987), regardless of whether differences are local, global, or both. In the conjunction search task, the difference between the target and the distractor is never smaller than it is in either the global search task or the local search task. It differs from half the distractors the same amount as in the local search task and half the distractors the same amount as in the global search task. If global and local features are part of the same description, then, the slope in the conjunction condition should be no more than the mean of the slope of the local and global conditions. Instead, it is much higher than either slope, indicating that the two features are distinct, and the visual system must integrate them together in a process that requires focal attention.

General Discussion

The goal of this research was to investigate the relations of shape coding of the global form of an object and descriptions of the local features from which the object's shape is composed. Logically, global shape must be computed from local elements, but a century of research in perception has shown that what is ultimately represented depends very little on the elements present during sensation (Koffka, 1935; Kanizsa, 1976; Navon, 1977; Tanaka, Kay,

Grinnell, Stansfield & Szechter, 1998; Pomerantz & Portillo, 2011; Baker & Kellman, 2018). At the same time, some information about local contour features is preserved after sensation (Erens, Kappers & Koenderink, 1993; Mamassian, Kersten & Knill, 1996). We aimed to clarify the degree to which local and global information are processed together and to better understand what information about a shape's local contour features is represented beyond the visual icon.

Our hypothesis was that the visual system handles local contour features and global shape descriptions are largely independently and in separate systems. This theory was partly inspired by past experiments studying the discrimination between circles and shapes formed by the addition of radial frequency patterns, which showed that detection of the target is different when high (local) and low (global) RF patterns are added to the contour (Jeffrey, Wang & Birch, 2002; Bell, Badcock, Wilson & Wilkinson, 2007). We also hypothesized that the visual system represents local contour features in a fundamentally different way than it represents global form. While representations of global form are highly descriptive about the curvature, relative size, and orientation of various parts, representations of local contour features do not specifically describe any of the individual elements. Instead, we proposed that the visual system estimates the distribution from which local contour features were sampled and that it is mainly sensitive to contour differences that change the properties of the distribution.

Differences in participants' sensitivity to global form and local contour features in Experiment 1 supported the idea that the two kinds of shape features are handled by different systems. Experiment 2 followed up on this finding by equating the amount of physical difference between locally and globally distinct shape pairs. If local and global information are processed in the same system, we expected the amount of physical difference between the pairs to be the main predictor for detectability of differences between the two shapes. Participants should therefore

have had similar sensitivity in the local and global condition when similarity was equated. To the contrary, sensitivity was markedly higher in shape pairs that differed in global features than shape pairs that differed in local features, implying that the two kinds of contour features are processed in different systems with different representational priorities.

The data in Experiment 1 also found that sensitivity was as high for a global shape difference as for a global and local shape difference. The shapes in the local and global condition always had strictly more total dissimilarity than the global only condition. If local and global contour features are processed in a single system, we would predict larger differences in the contour to correspond to better accuracy in detecting a shape change. A lack of any indication of additivity in local and global features suggests that they are handled in separate systems and do not necessarily interact in recognition tasks.

Participants' low sensitivity for phase shifted contour features in Experiment 3 also suggests independent local and global processing systems. In this condition, local contour features were preserved but shifted along the global form of the object. If participants represented local and global features together, then all the changes in the spatial relationships between the small and large features would result in a very different percept. Instead, the visual system's description of local contour features appears to be independent of their position relative to the global features of the shape.

Experiment 4 was an additional direct test of independence between local and global systems, using a converging method. Following Treisman and Gelade's (1980) Feature Integration Theory, we predicted that if local and global aspects of shape are distinct feature dimensions, then they should require focal attention to be integrated together. This hypothesis was confirmed in a visual search task in which the target could differ from distractors in local

statistical properties, global form, or a conjunction of the two. While targets that differed only in local statistical properties or global form popped out in the search array, targets with a conjunction of local and global features required serial search time. Targets' physical dissimilarity to distractors was just as high in the conjunction condition as in the local or global conditions, so a steeper search slope in the conjunction condition is evidence that the visual system needs to integrate to distinct features together to do the task.

The results of our experiments also support the idea that the visual system represents statistical properties of local contour features rather than each feature individually. In Experiments 1 and 2, local change trials inverted the polarity of all bumps along the object's contour. Even though no local contour feature was the same, participants had difficulty detecting a difference between locally different shape pairs. Participants' poor performance suggests that matching shapes based on local features is not primarily done by probability summation of local detectors (Loffler, Wilson & Wilkinson, 2003). Because all bumps are inverted, representation of even a small number of local contour features would provide high sensitivity to local changes in Experiments 1 and 2. On the other hand, if the visual system's primary tool for local feature comparison involves looking for a difference in the distribution from which features were sampled, sensitivity should be low for inverted bumps (assuming equal probability of positive and negative curvature bumps).

In Experiment 3, we proposed that the visual system primarily encodes the mean and variance of the amplitude of bumps and the frequency of bumps. We tested this by generating shape pairs that had matched frequency, amplitude, and variance. We compared sensitivity to these statistically matched pairs with shape pairs that had different frequency or different mean and variance of the amplitude. We also compared the statistically matched pairs with pairs whose

local contour features were the same but shifted slightly along the contour. Results confirmed that sensitivity was highest when the summary statistics for shapes' contour features differed in either frequency or amplitude.

Pop out of the target in the local condition in Experiment 4 lent further support for the hypothesis that the visual system represents statistical properties of local contour elements rather than individually representing each element. Despite all shapes in the array having unique local contour features, we found that participants quickly detected the target in the local condition, irrespective of the size of the array. Though they were not physically identical, all distractors had the same number of bumps along their contour, while the target always had 3.2 times more or fewer bumps. Pop out for local features appears to depend on different statistical properties of the local contour elements, not differences in the individual elements.

One aspect of our data that our hypothesis about statistical descriptions of local contour features does not explain is why sensitivity is greater than zero in statistically matched shape pairs. One possibility is that the visual system is sensitive to more kinds of statistical features than we equated in Experiments 1-3. Participants' higher sensitivity to the statistically matched shape pairs than the phase shifted shape pairs in Experiment 3 suggests that this might be the case. Given that even sensitivity to the inverted bumps in Experiments 1 and 2 is non-zero, though, it seems likely that the visual system can represent some local contour features with specificity, albeit rarely.

One possibility is that participants adopted a task-specific strategy for scrutinizing and memorizing a few local contour features and checking for their presence or absence in the second shape stimulus. This hypothesis is indirectly supported by the fact that sensitivity to local contour feature changes was lowest in Experiment 1, where shapes with high frequency contour

features could undergo a global change in addition to the local change. If memorizing a single local feature or small number of local contour features is effortful and unnatural, we would expect participants to do it less when other cues for discrimination are potentially available. A second explanation is that a local feature sometimes gets encoded as a global feature if, by random chance, it is particularly large or placed in a particularly salient position on the shape's global form, such as a local maximum or minimum. While either strategy could potentially lead to better-than-zero sensitivity to differences between statistically matched local contour features, neither is used often, as the observed sensitivity to local contour feature changes suggested that participants only truly detected a local difference in a small number (less than 10%) of trials.

How does the visual system abstract global form from high frequency contour features? One possibility is that the global processing system uses oriented detectors at different scales to extract the low frequency features from a shape. Multi-scaled detectors in early visual areas are sensitive to a particular orientation in visual field (Hubel & Wiesel, 1962; Gur, Kagan & Snodderly, 2005). Curvature along a contour could be estimated by the difference in preferred orientation of nearby detectors. We have proposed the existence of *arclets*, a higher-order neuron that is connected to two nearby oriented detectors and is maximally sensitive to a certain turn angle between the detectors (Garrigan, 2006; Kellman & Garrigan, 2007; Baker, Garrigan & Kellman, under review).

For a contour with no high frequency features, the most precise description of curvature would be obtained from arclets connected to two detectors of the smallest scale—as the size of each detector approaches a point, the difference between the curvature estimated by the turn angle and the contour's true curvature approaches zero. Still, arclets connected to larger-scale detectors have useful applications, including obtaining size invariance for contours scaled up or

down (Kellman & Garrigan, 2007). Another application of larger scale detectors might be to estimate the curvature of the contour's global shape in the presence of high frequency contour features, as illustrated in Figure 13, which shows a zoomed-in portion of a contour with added local contour features. Detectors at the finest scale are sensitive to orientation changes precipitated by local contour features, but larger scale detectors would remain sensitive to more global properties of the shape.

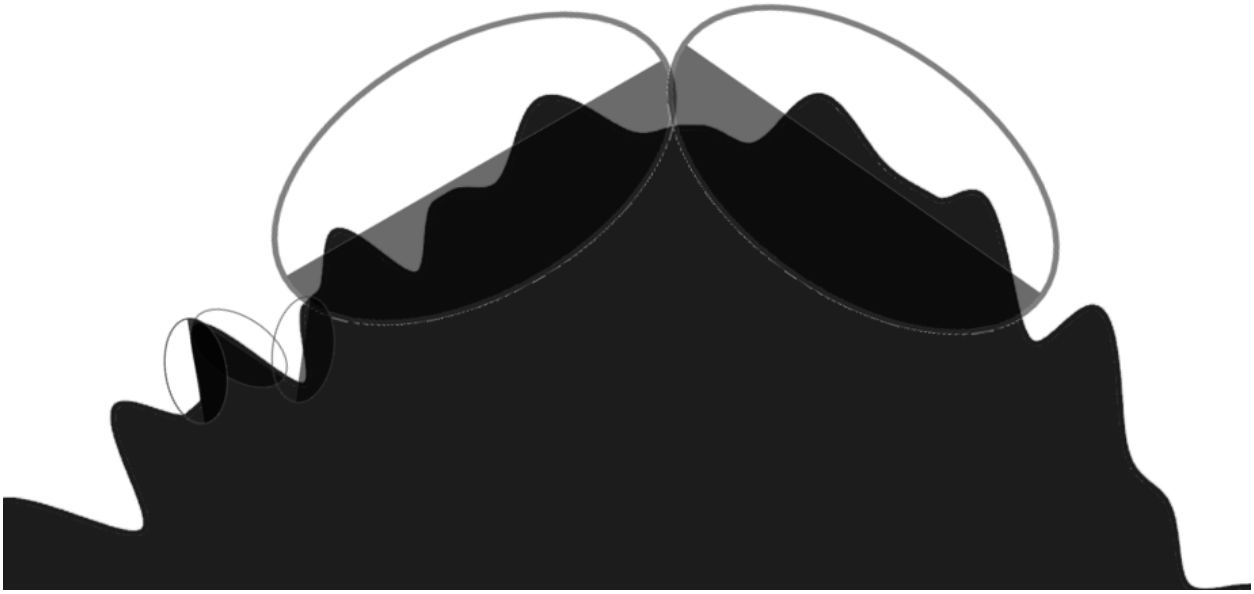

Figure 13. Comparison of two detectors' sensitivity to high frequency contour features. Three small detectors in the bottom left pick out changes along the contour from local features while the larger detectors on the top abstract over them.

It is unlikely that the visual system processes global information with detectors larger than some fixed threshold and local information with detectors smaller than it. Small contour features are often processed as part of an object's global shape when they are not accompanied by other similar small features. Likewise, we can still encode the global form of an object with high frequency contour features when it is scaled up, which would sometimes result in local elements crossing the threshold to be processed globally. A more likely hypothesis is that the

visual system settles on the scale of global detectors on a case-by-case basis depending on properties of the object's contour. One simple way it could do this is by looking at the profile of curvatures outputted by detectors at multiple scales. Going from the largest scale detectors to the smallest, there should be a point in which the curvature profile dramatically changes in contours with high frequency features. For the shape in Figure 13, any detector large enough to not be turned by the orientation of local elements will output a curvature profile corresponding to the underlying global shape of the object. However, at a certain point the scale of the detectors will be small enough that they are influenced by the orientation of local features, resulting in a curvature profile that includes dozens of turns and reverses in curvature polarity. The visual system might use the curvature profile outputted from arclets at the finest scale at which the description of the contour's curvature has a reasonable number of sign changes in forming an abstract representation of global shape. For now, this hypothesis is only a conjecture with some intriguing connections to other recent ideas in development about abstract shape representation (Baker, Kellman & Garrigan, under review; Baker, Kellman & Garrigan, in prep.). More specific future research is needed to develop a systematic theory of global shape extraction from multi-scale filters.

Conclusion

We conducted four experiments that point to a dissociation in human perception between local elements along a contour and the global form defined by relations between them. The system that encodes information about local elements does not precisely represent their properties individually, instead estimating a few statistical properties that are shared by local elements. The system that encodes global form represents parts of the object with much greater specificity and spatial precision. Although they are both concerned with representations of shape,

the local and global processing systems are distinct from each other and operate independently. Descriptions of the shape that require integration of local and global contour features can only be formed with focal attention as is needed for other distinct visual feature dimensions.

References

Arnheim, R. (1971). Art and entropy.

Baker, N. & Kellman, P. J. (2020). Constant curvature modeling of abstract shape representation. In preparation.

Baker, N., Garrigan, P., & Kellman, P.J. (2020). *A biologically inspired model of constant curvature representation of 2D contours. In preparation.*

Baker, N., Garrigan, P., & Kellman, P.J. (2020). Constant Curvature Segments as Building Blocks for 2D Shape Representation. Under review.

Baker, N., Garrigan, P., & Kellman, P.J. (2020). *Constant curvature segments as building blocks for 2D shape representation.* Submitted to *Journal of Experimental Psychology: General.*

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLoS computational biology*, *14*(12), e1006613.

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2020). Local features and global shape information in object classification by deep convolutional neural networks. *Vision Research*, *172*, 46-61.

Barenholtz, E., Cohen, E. H., Feldman, J., & Singh, M. (2003). Detection of change in shape: An advantage for concavities. *Cognition*, *89*(1), 1-9.

Bell, J., Badcock, D. R., Wilson, H., & Wilkinson, F. (2007). Detection of shape in radial frequency contours: Independence of local and global form information. *Vision Research*, *47*(11), 1518-1522.

Bell, J., Wilkinson, F., Wilson, H. R., Loffler, G., & Badcock, D. R. (2009). Radial frequency adaptation reveals interacting contour shape channels. *Vision Research*, *49*(18), 2306-2317.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive psychology*, *20*(1), 38-64.

Blum, H. (1973). Biological shape and visual science (Part I). *Journal of theoretical Biology*, *38*(2), 205-287.

Braddick, O. (1973). The masking of apparent motion in random-dot patterns. *Vision Research*, *13*(2), 355-369.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 679-698.

D'Zmura, M. (1991). Color in visual search. *Vision research*, *31*(6), 951-966.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of vision*, *11*(5), 14-14.

Erens, R. G., Kappers, A. M., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & psychophysics*, *54*(2), 145-156.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences, 103*(47), 18014-18019.

Garrigan, P. (2006). *Representation of contour shape*.

Gur, M., Kagan, I., & Snodderly, D. M. (2005). Orientation and direction selectivity of neurons in V1 of alert monkeys: functional relationships and laminar distributions. *Cerebral Cortex*, *15*(8), 1207-1221.

Hochberg, J., & Brooks, V. (1962). Pictorial recognition as an unlearned ability: A study of one child's performance. *The American Journal of Psychology*, *75*(4), 624-628.

Hoffman, D. D., & Richards, W. (1984). Parts of recognition, Cognition.

Jeffrey, B. G., Wang, Y. Z., & Birch, E. E. (2002). Circular contour frequency in shape discrimination. *Vision Research*, *42*(25), 2773-2779.

Kanizsa, G. (1976). Subjective contours. *Scientific American*, *234*(4), 48-53.

Kellman, P. J., & Garrigan, P. (2007). Segmentation, grouping, and shape: Some Hochbergian questions.

Kennedy, J. M. (1974). Perception, pictures, and the etcetera principle. *Perception: Essays in honor of James J. Gibson*, 209-226.

Koffka, K. (1999). Principles of gestalt psychology. 1935. *Lund Humphries, London*.

Loffler, G., Wilson, H. R., & Wilkinson, F. (2003). Local and global contributions to shape discrimination. *Vision Research*, *43*(5), 519-530.

Mamassian, P., Kersten, D., & Knill, D. C. (1996). Categorical local-shape perception. *Perception*, *25*(1), 95-107.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1517.

Navon, D. (1977). Forest before trees: the precedence effect of global features in visual perception. *Cognitive Psychol.*, *7*, 476-484.

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of a set-size effect. *Current directions in psychological science*, *4*(4), 118-123.

Pashler, H. (1987). Target-distractor discriminability in visual search. *Perception & Psychophysics*, *41*(4), 285-292.

Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic Gestalts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1331.

Quinlan, P. T., & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception & psychophysics*, *41*(5), 455-472.

Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception & Psychophysics*, *60*(6), 1067-1082.

Smithson, H., & Mollon, J. (2006). Do masks terminate the icon?. *Quarterly Journal of Experimental Psychology*, *59*(1), 150-160.

Tanaka, J. W., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition*, *5*(4), 479-496.

Treisman, A. (1993). The perception of features and objects in Attention: Selection. *Awareness and Control*.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97-136.

Vlaskamp, B. N., & Hooge, I. T. C. (2006). Crowding degrades saccadic search performance. *Vision research*, *46*(3), 417-425.

Wertheimer, M. (1923). Laws of organization in perceptual forms. *A source book of Gestalt Psychology*.

Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision research*, *38*(22), 3555-3568.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, *1*(2), 202-238.

Conclusion

Among the most fundamental questions in perception is how the visual forms abstract representations of shape. By abstract, we refer to representations that describe relations between parts, that are informationally economical, and that recode information so that constancy and similarity relations between shapes can be obtained. The first two papers put forward a model for how shape contours can be abstractly represented in a constant curvature framework. The third and fourth papers investigated situations which pose special challenges to abstract shape representation. In arrays of unconnected dots, the visual system encodes shape in the absence of a connected contour, imputing a contour based on the spatial relationships between elements. In shapes with local contour features, the visual system must obtain a global representation of shape that is invariant to contour noise. These cases offer unique insights and constraints into how representations of shape are formed in general. In both cases, features of the constant curvature model help to clarify how abstraction is ultimately accomplished. The reported studies span a wide range of stimuli and experimental paradigms. Participants were presented with novel contours, arrays of dots, and shapes built with random noise. Many experiments used the task of shape recognition in sequentially presented displays, but participants were also asked to give subjective reports, mentally rotate a stimulus, detect a target among random noise, and visually search in an array of distractors. Despite the diversity in stimulus types and perceptual tasks, the results of these chapters converge on a few general themes.

Constant Curvature Shape Representations. One theme that has emerged across these studies is the plausibility and empirical support for the use of constant curvature as a basic primitive for representing shape. In Chapter 1 and 2, we proposed a model for how a shape could be represented by constant curvature segments. The bounding contour of objects is very rarely

made up of regions of constant curvature. The task of the visual system, then, is not to detect regions of constant curvature along the physical contour of an object, but to organize various similar curvatures together and abstract them together into a single curvature value that approximates the curvature of that region. How contours are partitioned into constant curvature regions depends on the visual system's priorities for efficiency and precision. In a model with very little tolerance for curvature variation, shape representations will be built up from a great many small regions of constant curvature. They will have a high degree of physical resemblance to the contour they represent, but they will also be complex, straining visual memory capacities and lacking invariance to small contour variations. Models with too much tolerance for curvature variation will output representations of shape that have abstracted over contour features to which the visual system has sensitivity. In Chapter 1, we explored this tension between precision and simplicity using simple psychophysics experiments to estimate the visual system's tolerance for abstraction of contour properties. In Chapter 2, we showed how this model could be formed from initially subsymbolic inputs and adjusted the model so that partitioning was purely local and scale invariant.

Chapter 3 tested a behavioral prediction made by the constant curvature hypothesis about shape perception in dot arrays. According to structural information theory, straight line connections between dots are more economical than curvilinear arcs because arcs are a continuation of both length and angle, thus requiring twice as many bits of information as a straight edge (Smits & Vos, 1987, personal communication with Leeuwenberg). However, our hypothesis that shape representations are composed of constant curvature segments suggests that a single curved segment is more informationally compact than two straight segments joined at a corner. We predicted that dot arrays that signaled smooth corners would be more easily encoded

as shape than dot arrays that signaled angular corners. Across three experiments, a greater facility for encoding dot arrays with angles that give rise to smooth curves was confirmed.

Chapters 3 and 4 also contributed evidence to the multiscale filtering aspect of the constant curvature model proposed in Chapter 2. For both dot arrangements and contours that include local variations, the representation of shape that we ultimately encode appears to require integration over larger portions of the contour. In dot arrays, this is needed to get curvature from relations between spatially distant elements. In shapes with local contour features, the visual system must find the curvature of the shape's global form irrespective of local curvature variations. Analytical methods that find the curvature at every point along a contour's shape would be confounded by either stimulus type, but in our detector model framework, arclets exist at multiple scales, some of which will be large enough to get global shape representations with equivalence between a connected contour and 25 unconnected dots sampled therefrom or between two objects with the same overall shape but different local features.

From Subsymbolic to Symbolic Coding in Vision. Another point of intersection between all four papers is the need for theories of shape that connect subsymbolic outputs of early visual cortex with symbolic representations of the object. This was most explicitly developed in Chapter 2 in which we showed how a model of constant curvature representation could begin as subsymbolic activations and become fully symbolic. One of the critical insights from this work is that curvature, a mathematical notion, can be neurally instantiated from population codes of arclets that respond to activation levels from pairs of oriented detectors. Developing a biologically plausible model for the formation of shape representations also provided solutions to certain perceptual problems. As discussed above, the hypothesis that arclets exist at multiple scales just as oriented contrast detectors exist at multiple scales (Hubel &

Wiesel, 1962; Hubel & Wiesel, 1968) gives a way for the visual system to abstract over certain kinds of contour noise or impute abstract contour relations to arrangements of dots.

The results from Experiments 1 and 2 in Chapter 1 show how even simple contours undergo a recoding from subsymbolic to symbolic description in visual perception. In both experiments, participants were insensitive to physical differences along an open contour far beyond the stage when early neural circuitry could pick out a difference (Westheimer, 1976; Fahle, 1986). Although early foveal detectors are sensitive to all of the curvature differences tested in these experiments, the visual system only represents curvature differences that are large enough to be distinctly encoded in symbolic representation. The results of Experiment 3 also gave evidence that shapes are recoded into symbolic representations. We found a difference in how *perceptually* different two shapes were even when they were equally physically different and would thus have similarly different subsymbolic outputs.

In Chapter 3, the transition from subsymbolic to symbolic representations of shape is necessary to explain many of the perceptual capabilities participants demonstrated in processing dot arrays. For example, participants were tasked with matching two dot displays that were sampled from different positions along the same underlying shape. All local relationships between pairs or triplets of dots differed, as sampling points were independently jittered along the contour, but participants found equivalence between the two displays accurately, even across differences in orientation. Subsymbolic outputs from the two displays likely consisted of a few "key-points" detected by early visual mechanisms (Von Der Heydt, Peterhans & Baumgartner, 1984; Grossberg & Mingolla, 1985; Heitger, Rosenthaler, Von Der Heydt, Peterhans & Kubler, 1992) that differed from each other in almost every way. Only by extracting a symbolic code

from outputs of early visual areas could the visual system determine that the two dot arrays were matched in shape.

Symbolic description is also needed to explain perceptual performance on shapes made up of both local and global features in Chapter 4. Although early oriented detectors would signal edge properties for both kinds of contour features, people's sensitivity to local and global properties were very different. This suggests that global and local information are described with distinct symbolic codes. The local code appears to be a statistical description of local contour features rather than separate representation of individual features. Features processed in the global system appear to have much greater specificity.

Relations in Shape Perception. A third unifying theme that emerges from these studies is that representations of shape are descriptions of relations between elements rather than of elements themselves. In the models proposed in Chapters 1 and 2, partitioning depended not on the absolute orientation or position of edge fragments but on the angular relations between fragments—two parts of a contour were organized together if the angle between constituent elements was similar regardless of individual element features. Likewise, the angular and curved dot displays presented in Chapter 3 were matched in their physical properties such as the number of dots, the spacing between dots, and the properties of dots themselves, but participants were consistently better at perceiving and encoding shapes in displays where the dots signaled a curved contour. This suggests that the relations between dots affects people's perception of the contour far more than properties of individual features. The preeminence of relational information is also observed in Chapter 4. Participants showed little ability to detect a change in shape when all the local elements were inverted or resampled altogether so long as the relations between elements were preserved.

Future Directions

More remains to be done connecting models of constant curvature representation to shapes and configurations like the ones used in Chapters 3 and 4. Findings from those studies suggest that contour representation happens at multiple scales, which should be more fully incorporated into our model. In the case of dot configurations specifically, it would be interesting to develop a more fleshed out theory about how shape representation can be formed from only key-points. Beginning by connecting dot arrays with virtual (straight) lines (Wallach & O'Connell, 1953) and deriving perceptual curvature at the symbolic stage from the turn angle between these lines first direction.

Many questions about how the visual system extracts local and global contour features separately also remain to be answered. For example, what determines whether a contour feature is processed as local or global? To the degree that this depends on the amplitude of local features, is the threshold amplitude relative to the length of the contour more important or does retinal amplitude of a feature play a larger role?

Finally, how do the findings of these papers relate to object perception in 3D? The curvature of an object's projected contour will change when it is rotated in depth, due to foreshortening, so how does the visual system find equivalence between depth-rotated objects? More generally, is there a 3D shape description system that is built from surface patches, such that each patch is composed of two constant curvatures in orthogonal directions? (For example, the shape of a soda can along its sides is composed patches having constant curvature along cross sections of the can and zero curvature in the orthogonal direction.) Much of the effort of the reported chapters was aimed at connecting low level vision and middle level vision, but the

connection between middle vision and higher levels where the visual system obtains 3D structural descriptions of objects is equally important and interesting.

Conclusion

Shape is a crucial notion – mathematically, perceptually, and ecologically. It poses fascinating challenges at multiple levels for scientific study and explanation. The work described here advances our understanding of shape as abstract, as spanning early visual coding to symbolic representations, producing high-fidelity yet flexible representations that are central to perception, action, and thought.

References

Fahle, M. (1986). Curvature detection in the visual field and a possible physiological correlate. *Experimental Brain Research*, *63*(1), 113-124.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological review*, *92*(2), 173.

Heitger, F., Rosenthaler, L., Von Der Heydt, R., Peterhans, E., & Kübler, O. (1992). Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision research*, *32*(5), 963-981.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106-154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, *195*(1), 215-243.

Smits, J. T., & Vos, P. G. (1987). The perception of continuous curves in dot stimuli. *Perception*, *16*(1), 121-131.

Von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, *224*(4654), 1260-1262.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of experimental psychology*, *45*(4), 205.

Westheimer, G. (1976). Diffraction theory and visual hyperacuity. *American journal of optometry and physiological optics*, *53*(7), 362-364.