Skeletal representations of shape in human vision: Evidence for a pruned medial axis model

Vladislav Ayzenberg

Emory University, Psychology, Atlanta, GA, USA ✉

Yunxiao Chen

The London School of Economics and Political Science, London, UK

Sami R. Yousif

Yale University, Psychology, New Haven, CT, USA

Stella F. Lourenco

Emory University, Psychology, Atlanta, GA, USA ✉

A representation of shape that is low dimensional and stable across minor disruptions is critical for object recognition. Computer vision research suggests that such a representation can be supported by the medial axis—a computational model for extracting a shape's internal skeleton. However, few studies have shown evidence of medial axis processing in humans, and even fewer have examined how the medial axis is extracted in the presence of disruptive contours. Here, we tested whether human skeletal representations of shape reflect the medial axis transform (MAT), a computation sensitive to all available contours, or a pruned medial axis, which ignores contours that may be considered "noise." Across three experiments, participants ($N = 2062$) were shown complete, perturbed, or illusory two-dimensional shapes on a tablet computer and were asked to tap the shapes anywhere once. When directly compared with another viable model of shape perception (based on principal axes), participants' collective responses were better fit by the medial axis, and a direct test of boundary avoidance suggested that this result was not likely because of a task-specific cognitive strategy (Experiment 1). Moreover, participants' responses reflected a pruned computation in shapes with small or large internal or external perturbations (Experiment 2) and under conditions of illusory contours (Experiment 3). These findings extend previous work by suggesting that humans extract a relatively stable medial axis of shapes. A relatively stable skeletal representation, reflected by a pruned model, may be well equipped to support real-world shape perception and object recognition.

Introduction

The appearance of objects varies dramatically across viewpoints, and yet, humans recognize objects rapidly and accurately with seemingly little effort. Researchers have long known that object recognition begins by extracting the shape of the object (Biederman, 1987; Marr & Nishihara, 1978; Wagemans et al., 2008), but the underlying computations involved in shape perception remain poorly understood (for review, see Elder, 2018). One model that has garnered much attention posits that shape can be distilled down to its internal skeletal structure, specifically its medial axis. The medial axis is a summary representation of shape, which may facilitate object recognition by providing a low-dimensional set of properties that are stable across different viewpoints (Blum, 1967, 1973; Kimia, 2003). Although computer vision research has provided successful demonstrations of the medial axis model in object recognition (T.-L. Liu & Geiger, 1999; Sebastian, Klein, & Kimia, 2004; Trinh & Kimia, 2011), there are open questions about its biological plausibility, particularly how the human visual system computes the medial axes of shapes under variable viewing conditions. To fill this gap in the literature, we used a unique behavioral paradigm (described below as the "tap-the-shape" paradigm; Firestone & Scholl, 2014; see also Psotka, 1978) that attempts to capture how individuals represent the internal structure of two-dimensional (2-D) shapes. In addition to providing a replication of recent research, in which participants' responses were consistent with the medial axis of different 2-D shapes, we extended this work by determining how the medial axis is represented in cases of perturbation and illusory contours. As described below, we tested whether participants' responses in the tap-the-shape task reflect a skeleton consistent with the medial axis transform (MAT), a computation sensitive to every edge in a shape (Blum, 1967, 1973; Ogniewicz & Kübler, 1995),

or a pruned medial axis, which shows stability across contexts by ignoring edges considered "noise" (Blum & Nagel, 1978; Kimia, 2003; Shaked & Bruckstein, 1998).

The medial axis is classically defined as the set of all symmetry points within a shape having two or more closest points along an object's boundary (Blum, 1967, 1973). For most shapes, its structure is organized hierarchically such that there is typically a parent branch that describes the shape's global geometry as well as several secondary branches that "grow" off the parent branch and describe the local geometry (e.g., corners; Pizer, Oliver, & Bloomberg, 1987). In many contexts, the medial axis is a good summary representation because it represents the spatial configuration of a shape's contours with minimal information. The original implementation of the medial axis—the MAT—is a shape skeleton computed from all of the available contour information in the object. Under the MAT, the shape skeleton is altered drastically in response to any change to the shape. For example, even a subtle perturbation along the contour of an object (due to partial occlusion or damage) alters the medial axis in the form of additional branches. Consistent with such sensitivity in the visual system is evidence of medial axis processing in cortical regions known to be edge sensitive (i.e., V1; Hubel & Wiesel, 1959). More specifically, single-unit recordings with rhesus monkeys have revealed a heightened response to the medial axes of 2-D shapes from edge-detection neurons in V1 (Lee, 1996, 2003), and human participants have been found to display greater contrast sensitivity (a known proxy of activity in V1; Boynton, Demb, Glover, & Heeger, 1999) to Gabor patches as they neared the medial axis of the shape (Kovács, Fehér, & Julesz, 1998; Kovács & Julesz, 1994). Such findings suggest that the MAT formulation may support low-level shape processes such as figure-ground segmentation because it can specify all of the edges that are part of the object rather than the background (Ardila, Mihalas, von der Heydt, & Niebur, 2012; Li, 2000).

Yet, a long-standing criticism of the MAT is that the addition of axial branches in response to small shape disruptions may lead to unnecessarily complex skeletal structures. Such complexity could make object recognition difficult across contexts and with visually similar shapes (Feldman & Singh, 2006; Kimia, 2003; Shaked & Bruckstein, 1998). To address this concern, researchers have attempted to implement medial axis models that "prune away" extraneous branches. The purpose of pruning models is to reduce the complexity of the medial axis by extracting only those branches that succinctly describe an object's overall shape, rather than every local contour. Such descriptions are more stable across contexts (Attali, Boissonnat, & Edelsbrunner, 2009; T.-L. Liu & Geiger, 1999; Sebastian et al., 2004; Siddiqi, Shokoufandeh, Dickinson, & Zucker,

1999) and may be better able to support object recognition than the MAT (Feldman & Singh, 2006; Kimia, 2003; Shaked & Bruckstein, 1998). Consistent with such stability in human vision is research showing that participants are tolerant to various changes to an object's component parts (e.g., nonaccidental properties) as long as the overall skeletal structure remains intact (Ayzenberg & Lourenco, 2019; Lowet, Firestone, & Scholl, 2018). Moreover, the medial axis structure of three-dimensional objects has been decoded in the inferior temporal (IT) cortex (Hung, Carlson, & Connor, 2012; Lescroart & Biederman, 2013), an area known for its tolerance to border disruptions (Kourtzi & Kanwisher, 2001). In these studies, the decoding withstood a variety of contour changes that would be unlikely under a MAT computation. It has been suggested that evidence of medial axis sensitivity in early, edge-sensitive, visual areas may not reflect encoding of the medial axis within these areas but rather is the result of top-down feedback from IT (Hung et al., 2012; Kimia, 2003; Lee, 2003).

Computer vision studies of object recognition (e.g., Siddiqi et al., 1999), as well as neural and behavioral evidence of a medial axis computation (e.g., Hung et al., 2012), provide crucial support for the viability of skeletal representations in human vision. However, there remains debate about whether the human visual system extracts shape information exclusively in the form of the MAT computation or whether there is pruning of the medial axis. Recent work with human participants would seem consistent with the MAT in that participants' skeletal representations of 2-D shapes changed in response to perturbations along the shape's edge (Firestone & Scholl, 2014). Firestone and Scholl (2014) adapted a task developed by Psotka (1978) that involved participants tapping within an enclosed 2-D shape presented on a tablet computer. In this task, participants' responses reflected the structure of the medial axis in simple (e.g., rectangle and triangle) and complex shapes (e.g., "guitar" shape). Moreover, when participants were presented with shapes that included external border perturbations, their responses seemingly conformed to a skeleton with additional branches, consistent with the MAT, in which all of the contours were incorporated into the model.

Current study

In the current study, we used the tap-the-shape task to test whether medial axis representations were better described by a MAT or pruned model across various conditions of noise (Firestone & Scholl, 2014; Psotka, 1978). Although computer vision has long dissociated between MAT and pruned structures (Blum & Nagel,

1978; Shaked & Bruckstein, 1998), surprisingly little research has examined the biological plausibility of each. Determining whether individuals represent the medial axis according to a MAT or pruned formulation, especially under conditions of noise, is important for our understanding of shape perception, as well as the role that skeletal representations may play in object recognition.

As a first step toward filling this gap in the literature, we began by providing a replication of previous research that found that participants' responses reflected the medial axis of enclosed 2-D shapes on the tap-the-shape task (Experiment 1). To anticipate, we found that participants' collective responses conformed to the medial axis across different shapes (i.e., rectangle, T-shape, square, and arc) and that the medial axis was a better characterization of participants' responses than the shapes' principal axes, another viable model of shape perception (Marr & Nishihara, 1978; Sturz, Boyer, Magnotti, & Bodily, 2017). Moreover, we also demonstrated that these results could not be explained by an alternative, task-specific cognitive strategy that emphasized boundary avoidance.

In two subsequent experiments, we directly tested whether participants' responses in the tap-the-shape task were better characterized by an edge-sensitive MAT computation or an edge-tolerant pruning computation that removed extraneous medial axis branches in the presence of perturbations (Experiment 2) or illusory contours (Experiment 3). Under a MAT computation, such conditions would cause drastic alterations to the skeletal structure of a shape. That is, if the medial axis is computed from all of the available edges, including those that only minimally disrupt a shape, then participants' responses should reflect these disruptions with additional branches to the medial axis. By contrast, a medial axis model that incorporates pruning should be more robust to inconsistent edge information. That is, in the case of pruning, participants' responses would not reflect disruptions to the shape, such that changes to the medial axis should be minimized.

General methods

Each participant was presented with a single shape on a tablet computer and instructed to "tap the shape anywhere you like." Participants were first asked by an experimenter whether they wanted to participate in "an extremely short psychology experiment that will only take two seconds." If participants agreed, they were then presented with the shape on the tablet (oriented horizontally) and instructed to tap the shape. If participants hesitated to tap the shape, follow-up instructions were "just tap anywhere." Following a
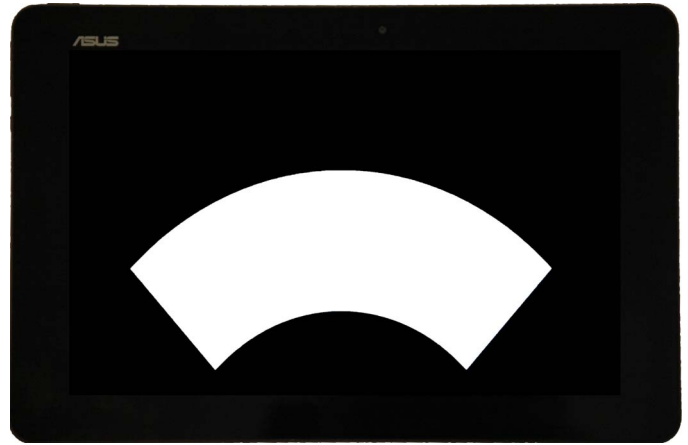

Figure 1. Cropped photograph of the tablet and stimulus display. In Experiments 1 and 2, each shape was presented as a white silhouette on a black background, as illustrated here. In Experiment 3, illusory shapes were presented using four black crescents on a white background (see Experiment 3). The location of the shape onscreen was randomized.

response, the shape disappeared (1,000 ms after response) and then reappeared for the next participant. Participants were given the opportunity to tap the shape only once and were immediately thanked following participation. We chose to collect a single response per participant for consistency with prior work (Firestone & Scholl, 2014) and to ensure independence of responses, which we reasoned would reduce the use of response strategies and could be crucial for accurately capturing the underlying perceptual representation (cf. Vul, Hanus, & Kanwisher, 2009; Vul & Pashler, 2008).

Each shape was tested one at a time until at least 200 responses were collected. Responses that fell outside the boundary of the shape were removed and replaced until 200 valid responses remained. Every effort was made to ensure that participants did not see the location of other participants' responses. The tablet was cleaned periodically so that fingerprints were not visible. All participants were adult pedestrians in public places of a metropolitan city. No demographic information was collected.

All stimuli were 2-D shapes presented as either white silhouettes on a black background (see Experiments 1 and 2; see Figure 1) or an illusory shape defined by four black crescents on a white background (see Experiment 3). We chose to present our shapes as silhouettes instead of outlines in Experiments 1 and 2 (cf. Firestone & Scholl, 2014) because the increased contrast made the shapes more clearly visible outdoors and because we wanted to ensure that perturbations were unambiguously perceived as a missing part of the shape rather than an imperfection on the screen. Our shapes were comparable in size to those used by Psotka (1978)

but larger than those used by Firestone and Scholl (2014). We chose to use larger shapes to ensure that participants had sufficient room to respond in all parts of the shape, even when perturbations were included. Shapes were presented on an Asus tablet computer with a capacitive 25.7-cm-diagonal touchscreen (1,200- × 800-px resolution; 21.7 × 14.6 cm) using a custom program written in Visual Basic (Microsoft). All shapes were presented in a random location onscreen.

General analyses

In all experiments, we first examined whether participants' collective responses differed from chance responding. More specifically, we tested whether their responses within each shape were closer to the model of interest (Experiment 1: medial and principal; Experiments 2 and 3: MAT and pruned) than would be predicted by chance. We then used latent variable analyses and goodness-of-fit metrics to compare the fit of different models to participants' responses. We also implemented the density ridge algorithm (Chen, Ho, Freeman, Genovese, & Wasserman, 2015) to better visualize the pattern of participants' responses and to examine how well they matched each model.

Comparisons to chance followed the procedure of Firestone and Scholl (2014). Using a Monte Carlo simulation, we generated 50,000 data sets of 200 randomly and uniformly sampled points for each shape. We then calculated the distance of each participant's response and each random point from the nearest point on the axis structure. Finally, we compared the mean distance of participants' responses to the mean distance of the best set of 200 random points from the simulation (i.e., the set closest to the axis). A mean response distance that is numerically smaller than the best set of simulated points for a shape suggests that the axis structure captured responses better than chance.

Latent variable models were implemented to determine the best fitting model of participants' responses (Experiment 1: medial vs. principal vs. boundary avoidance; Experiments 2 and 3: MAT vs. pruned). Comparing models exclusively on the basis of response distance is not appropriate because structures that occupy more area within a shape incur a statistical advantage.[1] By contrast, the latent variable models allowed us to estimate the fit of participants' responses to each structure while controlling for length (see the Appendix for the notation of the latent variable models). These analyses assume that if participants' responses correspond to the model of interest, then these responses should reflect a point along that model's structure and some tapping error (constrained by the border of the shape). Maximum likelihood estimation was used to determine the variance terms that best

described the distribution of participants' responses around each point of a model's respective structure (see the Appendix for the notation of the likelihood functions).

Participants' responses were analyzed with two latent variable models, each evaluated using a Bayesian information criterion (BIC; Schwarz, 1978). The first latent variable model assumed responses were sampled from a normal distribution with equal frequency and variance around every point for the structure of interest. The second model assumed responses were sampled with separate variance terms to the shape's center of mass and the remainder of the shape. This specific statistical model was tested to control for a known bias toward the centers of shapes (Huttenlocher & Lourenco, 2007; Melcher & Kowler, 1999; Vishwanath & Kowler, 2003), which has also been shown to influence responses in the tap-the-shape task (Firestone & Scholl, 2014). Moreover, we tested the center-controlled model using both a truncated normal distribution and, to account for response outliers, a t-distribution, which is tolerant to outliers (Casella & Berger, 2002). For each shape, we have displayed the results of the statistical model (equally distributed or center controlled) that produced the smallest BIC value (see Experiments 1–3; see also Supplementary Tables S1–S6 for results separated by latent variable model). As an estimate of uncertainty, we also display bootstrapped confidence intervals for the log-likelihood values. The model that produced the largest log-likelihood value and the smallest BIC value was taken as the best fitting model of participants' responses.

In a final analysis, we used a density ridge algorithm to visualize the cumulative structure of participants' responses (Chen et al., 2015). This data-driven approach determines the optimal smooth curves within a shape that pass through the centers of data clusters under a nonparametric statistical model. Regions with greater concentrations of points form mountain-like "ridges" upon which the curves are plotted. For each shape, smooth curves were estimated by running 15 iterations of the algorithm and by using a progressive variance parameter that ranged from 50 to 100 in steps of 20. The variance parameter described the density of the data (i.e., the expected "width" of the ridge) and provided a threshold for creating additional smooth curves. The density ridge algorithm allows for an unbiased method of estimating the collective structure of participants' responses.

Experiment 1

In a first experiment, we sought to replicate the findings of two previous studies (Firestone & Scholl,

		Comparison to chance		Model comparison			
Shape	Model	Response distance (SD)	Simulation distance	Log-likelihood	CI-low	CI-high	BIC
Rectangle	Medial	39.92 (39.22)	67.71	−1590.30[a]	−1611.66	−1562.81	3191.19
	Principal	76.36 (80.06)	80.06	−1626.50[b]	−1648.60	−1603.82	3263.59
	Boundary	—	—	−1607.44[b]	−1634.38	−1577.34	3225.47
T-shape	Medial	30.59 (26.76)	43.70	−1543.77[a]	−1563.08	−1517.73	3098.14
	Principal	59.17 (55.72)	67.44	−1575.05[a]	−1591.97	−1553.65	3160.70
	Boundary	—	—	−1562.23[b]	−1582.36	−1534.10	3135.06
Square	Medial	26.52 (24.62)	46.25	−1429.33[a]	−1451.51	−1392.63	2869.25
	Principal	61.48 (63.43)	63.43	−1451.65[b]	−1480.26	−1420.37	2913.90
	Boundary	—	—	−1435.21[b]	−1460.92	−1402.43	2881.01
Arc	Medial	32.57 (29.64)	52.58	−1497.68[a]	−1519.59	−1470.61	3005.96
	Principal	44.91 (48.59)	56.90	−1530.91[b]	−1552.22	−1501.40	3072.43
	Boundary	—	—	−1509.83[b]	−1535.52	−1479.34	3030.26

Table 1. Results for Experiment 1. *Notes*: Mean participant response and simulation distances, as well as goodness-of-fit metrics, are displayed for each shape and model (medial, principal, and boundary avoidance). A mean response distance smaller than the simulation distance suggests that the model outperformed chance. Log-likelihood values indicate how well each axis structure or boundary-avoidance model fit participants' data, and confidence intervals (95% CI) provide estimates of uncertainty. For each axis structure and boundary-avoidance model, we display the results of the statistical model (center-controlled model: normal or t-distribution) that produced the smallest BIC value. [a] Log-likelihood and CIs using a center-controlled model following a t-distribution. [b] Log-likelihood and CIs using a center-controlled model following a normal distribution.

2014; Psotka, 1978) in which it was found that participants' responses reflected the medial axis of enclosed 2-D shapes when they were simply instructed to place one point (with their finger, as in Firestone & Scholl, 2014; or a pencil, as in Psotka, 1978) within a shape. To provide a stronger test of medial axis extraction, we compared the medial axis to another biologically plausible model of shape perception, namely, the principal axis model (Marr & Nishihara, 1978). Although Psotka (1978) provided qualitative evidence in favor of the medial axis, he did not statistically compare the medial axis to other shape models. Firestone and Scholl (2014) compared a rectangle's medial axis to its diagonal axes and found that participants' responses were numerically closer to the medial axis. However, diagonal axes cannot be extracted in most shapes, and no model of shape perception (to our knowledge) emphasizes a shape's diagonal axes. Here, we compared a medial axis model to one based on the principal axes, which can be computed for any shape and which are thought to play a role in shape perception (Humphrey & Jolicoeur, 1988, 1993; Sturz et al., 2017; Warrington & James, 1986; Warrington & Taylor, 1973).

A model of shape perception based on principal axes suggests that humans extract shape via axes bisecting an object's center of mass (Chaisilprungraung, German, & McCloskey, 2019; Marr & Nishihara, 1978). More specifically, each shape is described by a major axis, which bisects the longest axis of the object, as well as minor axes, which bisect the shorter axes of the object (Marr & Nishihara,

1978; Warrington & James, 1986; Warrington & Taylor, 1973). Like the medial axis, the appeal of such a structure is that it provides a low-dimensional description of an object's overall shape. However, unlike the medial axis, the principal axes of a shape do not provide any description of the object's local geometry such as curvature or corners (Ambosta, Reichert, & Kelly, 2013; Cheng & Gallistel, 2005; Kelly & Durocher, 2011).

We also tested each axis structure against a boundary avoidance model to ensure that participants' responses were not a by-product of a task-specific cognitive strategy to simply tap *inside* the shape and, therefore, *away* from the boundaries. To address this possibility, we tested boundary avoidance models that were based on the following two assumptions. First, if participants' responses are systematically biased away from the boundary, then they would be distributed across a smaller portion of the shape's area whose edges are equidistant from the boundary. Second, if participants' responses do not correspond to any specific model of shape perception, their responses would be randomly distributed within that smaller region of the shape. Such a response pattern would be best fit by a uniform distribution—defined here as a grid of equally-spaced points inside the shape. Because it was unclear to what extent participants should avoid the boundary, we tested the boundary avoidance model with multiple possible degrees of avoidance.

In this experiment, we examined participants' responses in one of four enclosed 2-D shapes.

Participants were presented with either a rectangle (907 × 485 px; 16.5 × 8.8 cm), T-shape (638 × 682 px; 11.6 × 12.4 cm), square (480 × 480 px; 8.7 × 8.7 cm), or arc (convex hull: 989 × 470 px; 18 × 8.5 cm). The rectangle was chosen specifically to replicate previous studies that found that participants' responses were consistent with the medial axis of a rectangle (Firestone & Scholl, 2014; Psotka, 1978). We included a T-shape because it allowed us to test whether participants' responses conformed to the medial axis in a more complex, multipart shape. Moreover, overlap between the medial and principal axes within the rectangle and T-shape allowed for a strong test of specificity. If participants' responses conformed to one axis structure over the other within these shapes, despite their large overlap, then it would provide strong evidence for a given axis structure within the corresponding shape. As a different test, we included a square for which there was no overlap between the medial and principal axes. The advantage of this comparison was that it allowed for an unambiguous context to dissociate medial and principal axes. Finally, an arc was included as a test of generalization. The rectangle, T, and square shapes are exclusively composed of straight edges. The inclusion of an arc in this experiment allowed us to compare axial structures in the case of curvature.

See Figure 2 for illustrations of the medial and principal axes, as well as the best-performing boundary avoidance model for each shape tested in the current experiment. As described in the General Methods section, 200 valid responses were analyzed for each shape. Additional responses were excluded for falling outside the shape: one from the rectangle, two from the square, and three from the arc. No responses were excluded from the T-shape.

Results

See Figures 2 and 3 for the distribution of participants' responses within each shape and Table 1 for a summary of the response distances and model fit metrics.

Consistent with the findings of Firestone and Scholl (2014), we found that participants' responses were closer to the medial axes of each shape (rectangle, T, square, and arc) than were the best 200 points from the corresponding Monte Carlo simulations of 50,000 distributions (see Table 1). Participants' responses were also closer to the principal axes of each shape than the simulations (see Table 1). Thus, comparisons of participants' responses to the medial and principal axes within each shape revealed that both axis structures captured participants' responses better than would be predicted by chance.

As described in the General Analyses section above, participants' responses were also analyzed with latent variable models that accounted for response distance and axis length to determine whether participants' responses were better fit by a medial or principal axis model. As can be seen in Table 1, participants' responses were best fit by the medial axis in every shape, as indicated by the largest log-likelihood estimates and smallest BIC values. Thus, although both medial and principal axes captured participants' responses better than chance, the medial axis described participants' responses better than the principal axes of each shape (see Figure 2).

However, an alternative account of these results is that participants' responses instead reflected a boundary-avoidance strategy. More specifically, it is possible that their responses formed the medial axis as a by-product of a cognitive strategy to tap inside the white area on the tablet and away from the boundary of this area. To address this possibility, we fit participants' responses to boundary models with multiple scales of avoidance (grids comprising 20% to 100% of the shape's area in 20% steps) and then tested whether the best-performing boundary model (rectangle: 60%; T-shape: 60%; square: 80%; arc: 40%; see Supplementary Tables S4–S6 for the results of all boundary avoidance models) provided a better fit to participants' responses than medial or principal axes models (see Table 1). These analyses revealed that the boundary avoidance model provided a better fit to participants' responses than the shapes' principal axes (see Table 1). However, and crucially, the medial axis model was a better fit to participants' responses than the best-performing boundary model for each shape (see Table 1). These results suggest that participants' responses could not be accounted for by a task-specific strategy to tap inside the shape and away from the boundaries but, instead, that participants may have extracted the medial axes of the different shapes tested.

Finally, we examined the pattern of participants' responses created by the density ridge algorithm. As can be seen in Figure 3, the density ridges for the different shapes formed smooth curves that were most consistent with the medial axes of the shapes, providing converging evidence for the medial axis as the best description of participants' responses in different 2-D enclosed shapes. The density ridges were inconsistent with the principal axis model and further rule out the possibility of boundary avoidance as an explanation of participants' responses.

Discussion

The results of this first experiment provide a replication and extension of prior research (Firestone &
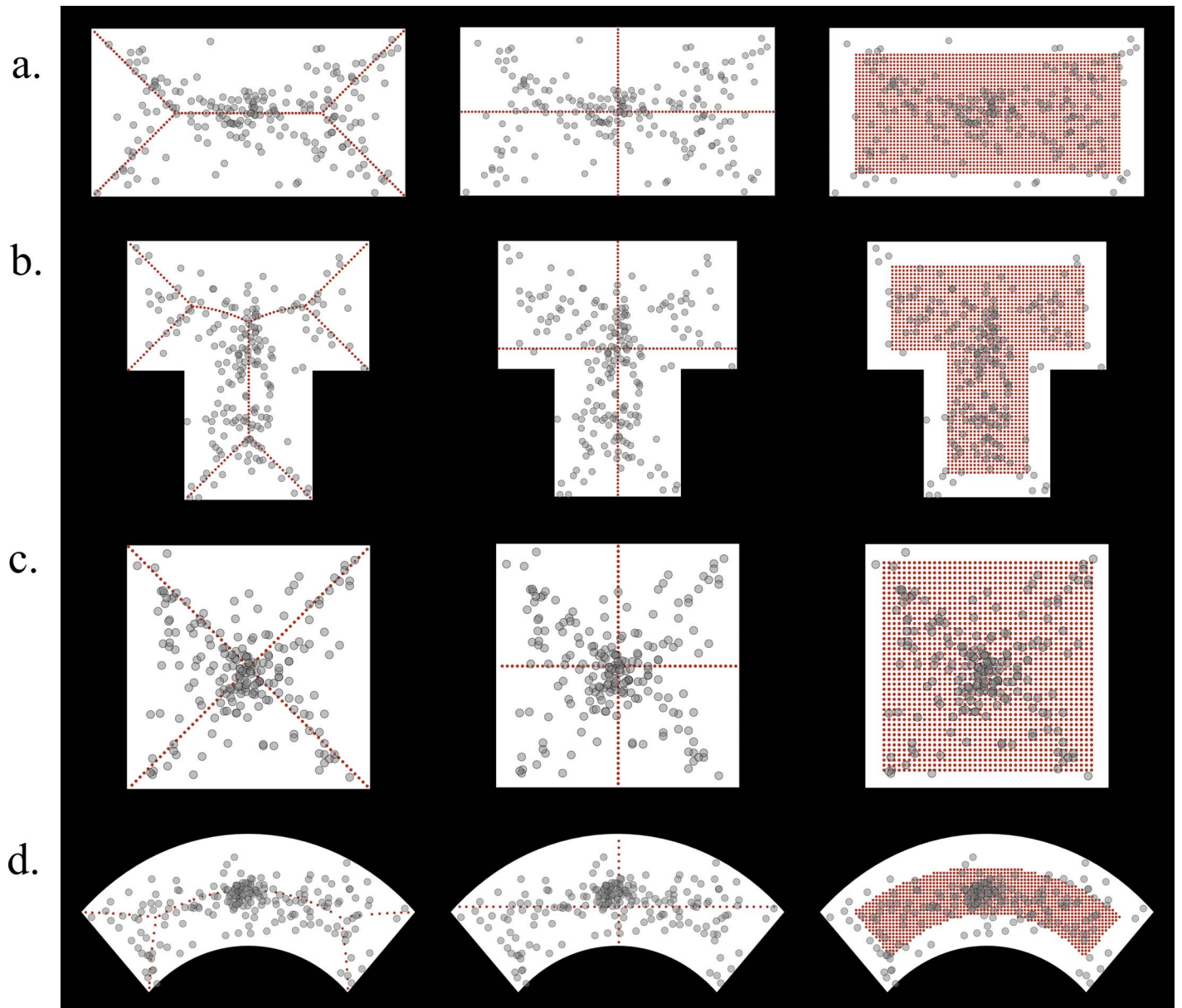
Figure 2. The different shapes used in Experiment 1: (a) rectangle, (b) T, (c) square, and (d) arc. Gray circles represent individual responses. Participants' responses for each shape are presented separately against the medial (left column, red dashed lines) and principal axes (middle column, red dashed lines), as well as the best-performing boundary-avoidance model (right column, red grid). Shapes are presented against a black background to mirror their presentation to participants on the tablet. Shapes are not drawn to scale.

Scholl, 2014; Psotka, 1978). We found that the medial axis model best fit participants' responses across all of the shapes tested. The medial axis model characterized the pattern of participants' responses better than the principal axis model, even in shapes with a high degree of overlap between models, such as the rectangle and T-shape. Moreover, the square provided a clear case of a dissociation between medial and principal axes, and the arc provided an example of generalization beyond straight-edged shapes. Importantly, these results could not be explained by a task-specific cognitive strategy to simply tap the white area on the tablet by avoiding the boundaries because participants' responses fit the medial axis better than the best-performing boundary avoidance model for each shape.

Although principal axes have been proposed as a viable summary representation of shape (Chaisilprun-graung et al., 2019; Marr & Nishihara, 1978; Sturz et al., 2017), our data do not provide support for extraction of principal axes of 2-D shapes. In this experiment, the principal axis model was the worst-performing model in all of the shapes tested—demonstrating a worse fit than the medial axis model and even against our control for boundary avoidance.
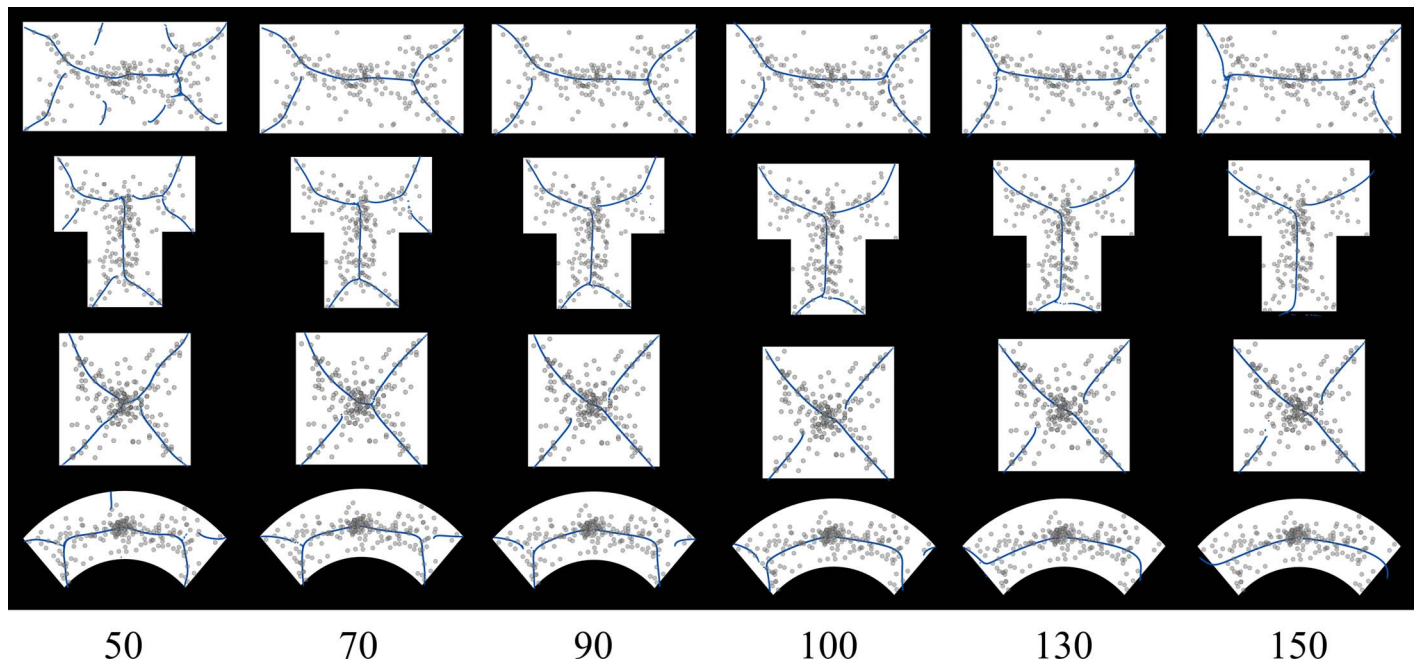
Figure 3. Smooth curves from the density ridge algorithm with increasing variance parameters. Each row displays a single shape from Experiment 1. Each column displays the smooth curves from each variance parameter (variance parameters displayed along the bottom). Shapes are presented against a black background to mirror their presentation to participants on the tablet. Shapes are not drawn to scale.

The above-chance performance of the principal axis model likely reflected the fact that it bisects perceptually prioritized areas, such as the center (center not controlled for in chance comparisons; Huttenlocher & Lourenco, 2007; Melcher & Kowler, 1999; Vishwanath & Kowler, 2003), or perhaps because it overlaps with the medial axis in some shapes.

As already noted, an especially important comparison in this experiment was between the medial axis model and a cognitive strategy based on boundary avoidance. This comparison was crucial because the medial axis is, by definition, the collection of points within the shape, equidistant from at least two points on the boundary. Accordingly, much of an object's medial axis structure occupies areas that are maximally distant from the borders of the shape such that responses in these regions are also consistent with a boundary avoidance strategy. However, the medial axes of the shapes tested here also include branches that are inconsistent with a boundary avoidance strategy, such as those that extend into the corners of the shape. If participants' responses reflected a strategy to tap "inside" and "away" from boundaries, then the more parsimonious prediction would have been that the location of their responses should have been random, once they were a safe distance away from the shape's border. Yet the latent variable analyses suggest that participants' responses were best fit by the medial axes of the shapes, not a boundary avoidance model. Moreover, most iterations of the density ridge algo-

rithm suggest that participants' responses clustered along the entire medial axis structure, including branches to the corners. These results are consistent with the findings of Firestone and Scholl (2014), who found that participants' responses adhered to the medial axis of a perceived shape, not the prescribed response area. Although it is possible that other unknown strategies could be responsible for participants' responses with these shapes, we have ruled out an obvious nonperceptual account of the data on this task. Thus, to the extent that the tap-the-shape task reveals the structure of human shape representations, our data are consistent with extraction of the medial axis during 2-D shape perception (see the General Discussion section for further discussion).

Experiment 2

Having found that participants' responses were most consistent with the medial axes of different 2-D shapes, we next tested which formulation of the medial axis was best characterized by participants' responses under conditions of perturbation. More specifically, we tested whether participants' responses were better described by the MAT formulation or a medial axis that incorporates pruning (Shaked & Bruckstein, 1998). The MAT predicts a medial axis that accommodates every available contour of a shape, such that perturbations,

regardless of size, lead to the growth of new axial branches. By contrast, a pruned computation predicts that the degree of medial axis accommodation will be proportional to the degree of change induced by the perturbation, thereby allowing for greater stability across contexts (Kimia, Tannenbaum, & Zucker, 1995; Shaked & Bruckstein, 1998). Importantly, the goal of the current experiment was not only to provide an answer to the general question of whether shape skeletons in human vision are better described by models that incorporate pruning but also to characterize the degree of pruning by the perceptual system. Although some models of pruning remove only those branches that describe the perturbation, leaving the remaining skeletal structure intact (e.g., Giblin & Kimia, 2003; H. Liu, Wu, Zhang, & Hsu, 2013), other models are more stringent such that they also remove branches that describe other aspects of the local geometry such as the corners of the shape (e.g., Ebert, Brunet, & Navazo, 2002; Feldman & Singh, 2006; Telea, Sminchisescu, & Dickinson, 2004). Because of the diversity of algorithms in the literature, our pruning models were created to exemplify two general classes of models (cf. Attali et al., 2009; Wieser, Seidl, & Zeppelzauer, 2017). More specifically, we tested a lenient pruning model that included branches describing the local geometry and a stringent model without these branches. Both types of pruning are consistent with a hierarchical organization of the medial axis. However, whereas parent branches (from which other branches grow) and those describing larger portions of the shape (e.g., branches to the corners) are less likely to be pruned by a lenient model, both types of branches are pruned by a stringent model. The lenient pruning model was defined by first computing the skeletal structure according to the MAT and then removing the new branches elicited by the perturbation (those lowest in the hierarchy). For the stringent pruning model, we further removed branches describing the local geometry (i.e., the outer corners).

Firestone and Scholl (2014) found that participants' touches within enclosed 2-D shapes conformed to a medial axis structure that accommodated an external border perturbation, suggesting a MAT computation. More specifically, they found that participants' responses reflected a medial axis with additional branches to the perturbation. However, they tested external perturbations of only one size, leaving it unclear whether participants would extract additional branches in other conditions where the perturbation was smaller or when the perturbation was placed internally.

Psotka (1978) found that the presence of perturbations affected the placement of participants' responses within circles, but the pattern did not conform to any known model of shape. More specifically, he found that participants largely avoided responding near small dot-

like internal perturbations, except when the perturbations were placed on the circle's medial axis (i.e., a point in the center). It is difficult, however, to compare MAT and pruned models for a circle because few algorithms (to our knowledge) are able to accurately calculate its medial axis when internal perturbations are introduced. Therefore, it is not clear what MAT and pruned medial axes should look like in a circle with internal perturbations.

To better understand medial axis extraction across a variety of perturbation conditions, in the current experiment we included externally and internally placed perturbations. Importantly, we tested participants with rectangular shapes that allowed us to compare MAT and pruned models. If the MAT computation is the better description of human skeletal representations, then participants should extract additional branches regardless of perturbation size or placement, because the MAT is computed from all of the available edges. By contrast, if a pruned medial axis structure provides a better characterization of skeletal representations, then no additional branches should be observed with smaller perturbations or perturbations placed internally because they are minimally disruptive and the added complexity of new branches in these cases does not provide an improved shape description (Feldman & Singh, 2006). It was less clear, however, how much pruning there would be with the large external perturbations. One possibility was that these perturbations would elicit new branches as in Firestone and Scholl (2014), with no pruning. Another possibility was that there would be no additional branches, reflective of pruning. Furthermore, to provide a better characterization of how participants extract the medial axis in the tap-the-shape task, we tested multiple degrees of pruning for every condition.

Participants were tested in one of four conditions: a rectangle (907 × 485 px; 16.5 × 8.8 cm) with a small (225 × 15 px; 4.1 cm × 0.3 cm) or large (225 × 125 px; 4.1 cm × 2.3 cm) external perturbation along the shape's border (see Figure 4a,b); a rectangle (931 × 412 px; 16.9 × 7.5 cm) with a small (22 × 22 px; 0.4 × 0.4 cm) or large (225 × 125 px; 4.1 × 2.3 cm) internal perturbation (see Figure 4c,d). Perturbations were meant to simulate natural shapes with disrupted contours as a result of damage, deletion, or partial occlusion by another object (Leyton, 1989).

As described previously, 200 valid responses were analyzed in each condition. Additional responses were excluded for falling outside the boundaries of the shape: two from the condition with the small external perturbation, nine from the condition with the large external perturbation, seven from the condition with the small internal perturbation, and six from the condition with the large internal perturbation. Of the excluded responses, 10 fell inside the perturbations (two in the large external perturbation condition, two in the
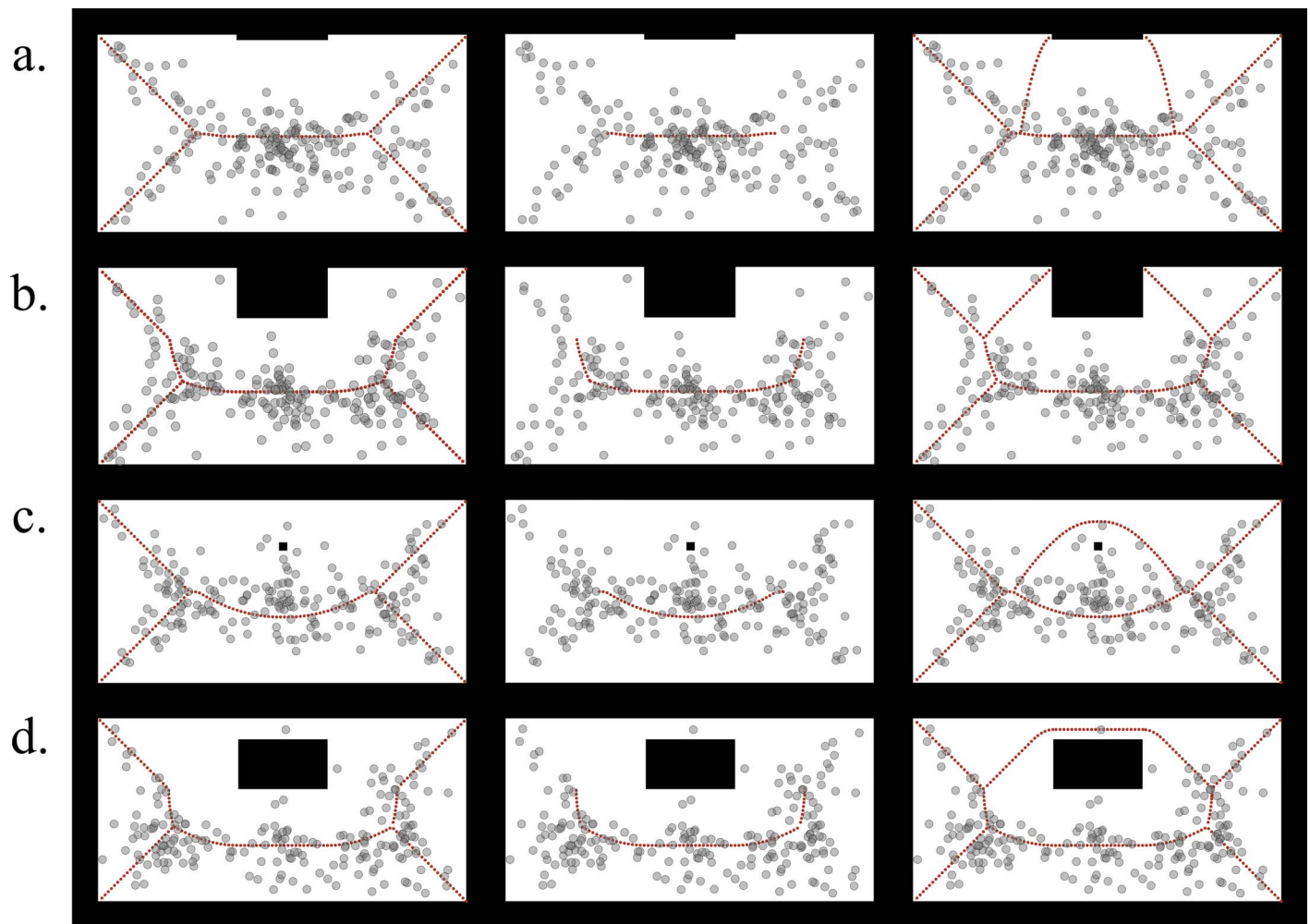
Figure 4. The four conditions from Experiment 2: rectangle with (a) small and (b) large external perturbations, and rectangle with (c) small and (d) large internal perturbations. Gray circles represent individual responses. Participants' responses are presented against a medial axis with lenient pruning (left column, red dashed lines), a medial axis with stringent pruning (middle column, red dashed lines), and the MAT structure (right column, red dashed lines). Shapes are presented against a black background to mirror their presentation to participants on the tablet. Shapes are not drawn to scale.

small internal perturbation condition, and six in the large internal perturbation condition).

Results

See Figures 4 and 5 for the distribution of responses within each shape and Table 2 for a summary of the response distances and model fit metrics.

Comparisons with the best set of 200 points from the Monte Carlo simulation revealed that participants' responses in all conditions (small and large, external and internal perturbations) were closer to the axes of the MAT than would be predicted by chance (see Table 2). Likewise, participants' responses in all conditions were closer to both pruned medial axis models than the best set of 200 points from the simulation (see Table 2). These analyses demonstrate that both the MAT and

pruned medial axes capture participants' responses better than chance but do not distinguish between the axis structures of interest.

Next, participants' responses were analyzed with the latent variable models and compared on the basis of their fit to each axis structure. As can be seen in Table 2, in all conditions, participants' responses were best fit by lenient pruning of the medial axis, as indicated by larger log-likelihood estimates and smaller BIC values. Moreover, the density ridge algorithm generated smooth curves that were most consistent with lenient pruning in all conditions, as indicated by the general presence of branches describing the shapes' corners but not the perturbations (see Figure 5). Taken together, the analyses suggest that a lenient pruning model best characterizes participants' responses in conditions of perturbation for both sizes and in both external and internal contexts.

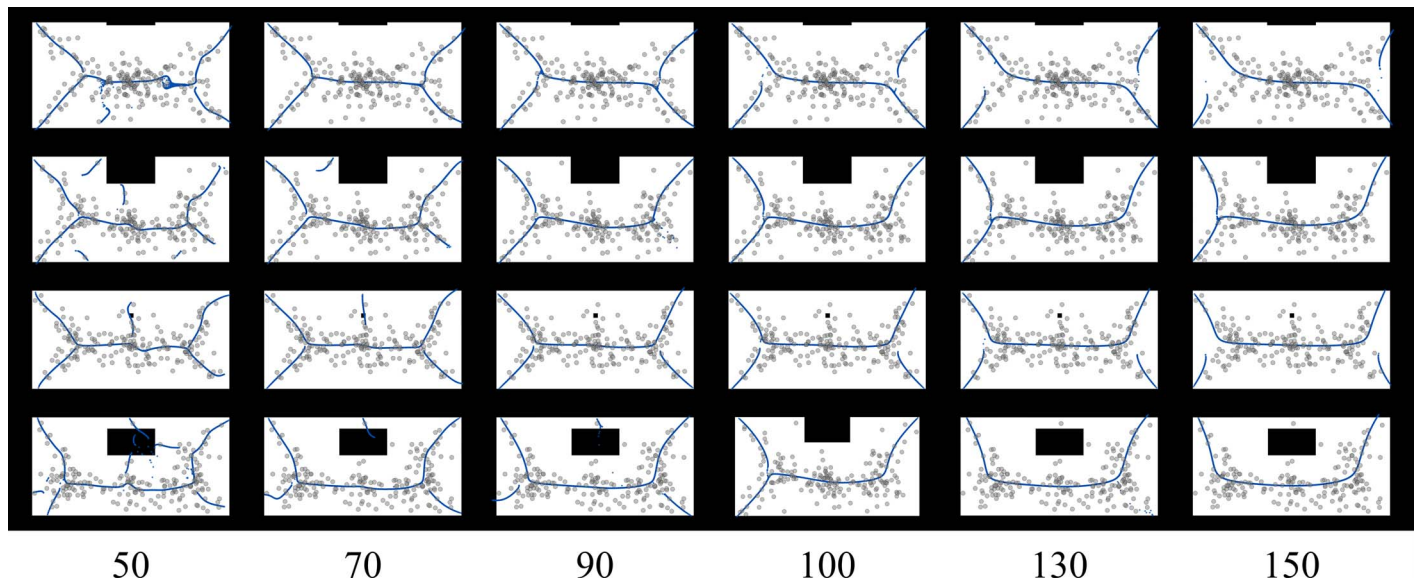50   70   90   100   130   150

Figure 5. Smooth curves from the density ridge algorithm with increasing variance parameters. Each row displays a single shape from Experiment 2. Each column displays the smooth curves from each variance parameter (variance parameters displayed along the bottom). Shapes are presented against a black background to mirror their presentation to participants on the tablet. Shapes are not drawn to scale.

Discussion

Our results show that participants' responses were best fit by a leniently pruned medial axis model in every condition tested, not the MAT or a stringently pruned model. Participants' responses did not reflect new branches in any perturbation condition, but responses did reflect branches to the outer corners. Thus, participants' responses were consistent with a medial axis representation that is robust to noisy edges but retains a sensitivity to important aspects of the local geometry. Such a representation creates a medial axis that may be more stable across minor disruptions to the shape while retaining enough information to

		Comparison to chance		Model comparison			
Shape	Model	Response distance (SD)	Simulation distance	Log-likelihood	CI-low	CI-high	BIC
Small external	L. Prune	40.88 (36.87)	66.4	−1559.70[a]	−1583.37	−1530.27	3129.99
	S. Prune	90.19 (77.68)	131.22	−1594.53[a]	−1618.83	−1564.93	3199.65
	MAT	35.69 (34.11)	52.27	−1573.62[a]	−1600.10	−1542.85	3157.84
Large external	L. Prune	35.49 (32.69)	60.82	−1577.30[a]	−1595.00	−1553.96	3165.19
	S. Prune	60.21 (54.66)	95.54	−1603.09[b]	−1628.29	−1574.79	3211.47
	MAT	34.29 (30.54)	49.25	−1598.63[a]	−1614.15	−1577.70	3207.86
Small internal	L. Prune	38.55 (65.23)	65.23	−1558.91[a]	−1577.33	−1536.19	3128.42
	S. Prune	82.91 (63.17)	122.83	−1631.18[c]	−1651.33	−1600.88	3272.96
	MAT	33.43 (24.10)	46.65	−1580.35[a]	−1597.07	−1558.50	3171.30
Large internal	L. Prune	36.01 (34.44)	60.11	−1589.99[c]	−1605.18	−1568.17	3190.57
	S. Prune	62.01 (50.99)	90.61	−1599.86[c]	−1618.97	−1575.33	3210.32
	MAT	33.19 (28.00)	41.56	−1635.52[a]	−1646.52	−1612.60	3281.63

Table 2. Results for Experiment 2. *Notes*: Mean participant response and simulation distances, as well as goodness-of-fit metrics, are displayed for each shape and axis structure (L. Prune = lenient pruning, S. Prune = stringent pruning, and MAT). A mean response distance smaller than the simulation distance suggests that the model outperformed chance. Log-likelihood values indicate how well each axis structure fit participants' data, and confidence intervals (95% CI) provide estimates of uncertainty. For each axis structure, we display the results of the statistical model (center-controlled model: normal or t-distribution; or equal distribution model: normal distribution) that produced the smallest BIC value. [a] Log-likelihood and CIs using a center-controlled model following a t-distribution. [b] Log-likelihood and CIs using an equal distribution model following a normal distribution. [c] Log-likelihood and CIs using a center-controlled model following a normal distribution.

recover the shape's unperturbed structure. These characteristics may be especially important for object recognition in cases where the object is partially occluded or where there is damage to the object's contour (Montero & Lang, 2012; Shaked & Bruckstein, 1998).

Although perturbations did not elicit new medial axis branches, they did cause some curvature to the existing branches. It is possible that curvature, of the parent branch in particular, may be an informative descriptor of the object's global shape and causal history (Leyton, 1992; Spröte & Fleming, 2016; Spröte, Schmidt, & Fleming, 2016). Indeed, even stringent pruning models appear to incorporate some curvature of the parent branch in the presence of a perturbation. Nevertheless, because the focus in the current study was on pruning of excess branches, we did not systematically assess the extent of curvature under these conditions. Importantly, these findings suggest that, even if skeletal representations show some accommodation to the perturbations, this accommodation does not increase the complexity of the medial axis in the form of new branches.

Given previous findings (i.e., Firestone & Scholl, 2014; Psotka, 1978), one might have expected that the perturbations in this experiment would have affected participants' responses to a greater degree. Firestone and Scholl (2014) found that, in the presence of an external perturbation, participants' responses appeared to conform to the MAT rather than the medial axis of an unperturbed shape. Similarly, Psotka (1978) found that the presence of internal perturbations influenced participants' responses within a circle, although they did not correspond to any known model of shape. By contrast, we found that perturbations had relatively little impact on participants' responses and were best fit by a leniently pruned medial axis, not the MAT. What might account for the differences between studies?

The first possibility is that the results found by Firestone and Scholl (2014) reflect the statistical advantage of the MAT under conditions of external perturbations. More specifically, Firestone and Scholl (2014) compared models on the basis of response distance in which longer axis structures incur a statistical advantage. Firestone and Scholl (2014) did equate for the number of axis points in each model, but this method may not have fully controlled for length.[1] In the current study, we compared models on the basis of their fit to participants' responses, such that the explanatory power of each axis point was considered in combination with response distance (see the General Analyses section and Appendix). When the length of each model is controlled for using this method, we find that participants' responses are best fit by a leniently pruned medial axis model with branches to the outer corners but not to the perturbations.

It is difficult to compare the results of our internal perturbation conditions to those of Psotka (1978) because, in his study, participants' responses did not correspond to any known model of shape, and he did not statistically compare medial axis models. Indeed, as mentioned previously, it is especially difficult to compare MAT and pruned models in a circle containing internal perturbations. Nevertheless, one might wonder why internal perturbations appeared to have had a greater effect on participants' responses in Psotka's (1978) study than they did here. It is possible that the simple medial axis of a circle (a point in the center) has a weaker influence on tapping behavior and therefore may be more likely to be overshadowed by other cognitive processes. Indeed, Psotka (1978) speculated that participants' responses in the perturbed circle conditions consisted of two processes: (a) medial axis extraction and (b) conscious effort by participants to "balance" the dot-like perturbations with their own responses to create a more symmetrical space within the figure. Here, the effect of these cognitive processes may be reduced because we used shapes with more extensive medial axes. Indeed, using rectangles, which have clearly distinguishable and extensive MAT and pruned medial axes, we found that responses were most consistent with a leniently pruned model for both internal and external perturbations conditions.

Another notable difference across studies is the use of shape silhouettes rather than outlines as in Firestone and Scholl (2014) and Psotka (1978). Could this difference in stimulus presentation have affected participants' responses? In the case of external perturbations, one possibility is that the use of silhouettes might have discouraged participants from responding near the edge of the shape, including the perturbation. Here, the black background could have created a region that participants were less likely to tap, causing them to respond farther from the perturbation than they would have otherwise, thereby leading to a pruned medial axis result. Although we cannot rule out this possibility directly, we suggest that it is unlikely. More specifically, if participants avoided responding near the boundary of the shape, then they should have also avoided responding near the outer corners, which would have created a response pattern most consistent with a stringently pruned model without branches to the corners. However, we found that the stringently pruned model provided the worst fit to participants' responses in most of the shapes tested. Moreover, the density ridge algorithm consistently created smooth curves into the outer corners of the shape, suggesting that participants did not avoid these areas.

In the case of internal perturbations, the use of silhouettes may have caused the perturbations to be seen as deletions (i.e., holes in the shape) rather than as partial occlusions. Deletions and occlusions can cause

different shape percepts (Bregman, 1981; Wagemans et al., 2012), and this perceptual difference could have led to a different pattern of responding in the tap-the-shape task. However, if anything, a shape deletion should have been more likely to create additional branches (and thus be more consistent with the MAT) than a partial occlusion. More specifically, symmetrical shape deletions, like the ones used in the current experiment, could be perceived as a permanent aspect of the shape (Spröte et al., 2016) and therefore more likely to be incorporated into participants' medial axis representations. That we found evidence of medial axis pruning in this case may suggest that pruning algorithms in humans are especially robust to perturbations. Nevertheless, more research will be needed to better understand how perturbation type (i.e., internal versus external) and stimulus presentation (i.e., silhouettes versus outlines) may influence participants' responses in the tap-the-shape task (see the General Discussion section for further discussion).

Experiment 3

In a third experiment, we asked whether physically present contours are necessary to elicit the medial axis. Classic work on illusory shapes suggests that shape perception can take place without real contours (Kanizsa, 1976). Perhaps the most famous examples are Kanizsa shapes, in which it is found that individuals infer a complete shape from crescent inducers specifying the corners. In this case, the crescents form the corners of a shape, inducing a percept of an actual, complete shape. Thus, if participants incorporate illusory contours in the tap-the-shape task, then the prediction is that their responses should reflect the medial axis of a complete shape. However, illusory contours are another point of weakness for the MAT computation, which creates a medial axis with branches that extend toward the missing contours of the shape (Johannes, Sebastian, Tek, & Kimia, 2001; Kimia, 2003; see Figure 6). By contrast, a pruned medial axis, according to the lenient formulation found in Experiment 2, would not include branches extending toward the outside of the shape but instead would generate a skeleton consistent with a complete shape (see Figure 6). Thus, in the current experiment, we examined whether participants' responses reflected the medial axis in a shape defined by illusory contours, as well as whether their responses were better fit by the MAT or pruned formulation. This question is not only important for understanding medial axis representations under different conditions but also allows us to further evaluate the validity of the tap-the-shape task as a measure of shape perception. That is, if participants'

responses in Kanizsa shapes reflect the medial axis of a complete shape, this would provide further evidence that the tap-the-shape task incorporates participants' perception of shape rather than a response strategy exclusively.

Participants were tested with two Kanizsa shapes— rectangle (illusory shape: 907 × 485 px; 16.5 × 8.8 cm) and square (illusory shape: 480 × 480 px; 8.7 × 8.7 cm)—based on Experiment 1. The perceived shape was defined by four black crescents presented on a white background (see Figures 6 and 7). Although the MAT predicts that the axial branches extend indefinitely outside the shape, we limited our analyses to the branches and responses that fell within the area of the perceived shape. There were 13 excluded responses in the Kanizsa rectangle and 19 for the Kanizsa square; many of the excluded responses fell inside the crescents, not outside the perceived shape (11 in the rectangle condition and 9 in the square condition).

Results

See Figures 6 and 7 for the distribution of responses within each shape and Table 3 for a summary of the response distances and model fit metrics.

Comparisons with the best set of 200 points from the Monte Carlo simulation revealed that participants' responses were closer to the axes of the MAT formulation for both the Kanizsa rectangle and square (see Table 3). Likewise, participants' responses were closer to a pruned medial axis than the best set of 200 points in both shapes (see Table 3). Thus, as in our previous experiments, these analyses demonstrate that different axis structures capture participants' responses better than chance, but they do not distinguish between them.

Next, participants' responses were analyzed with the latent variable models and compared on the basis of their fit to each axis structure. As can be seen in Table 3, in both conditions, participants' responses were best fit by a lenient pruning model in both the Kanizsa rectangle and square, as indicated by larger log-likelihood estimates and smaller BIC values. Moreover, the density ridge algorithm generated smooth curves that were comparable to those of a pruned medial axis (see Figure 7).

Discussion

The results of Experiment 3 suggest that participants' responses reflect the medial axis of inferred shapes defined by illusory contours (Kanizsa, 1976; Wagemans et al., 2012), as we previously found for real, complete 2-D shapes (Experiments 1 and 2). The
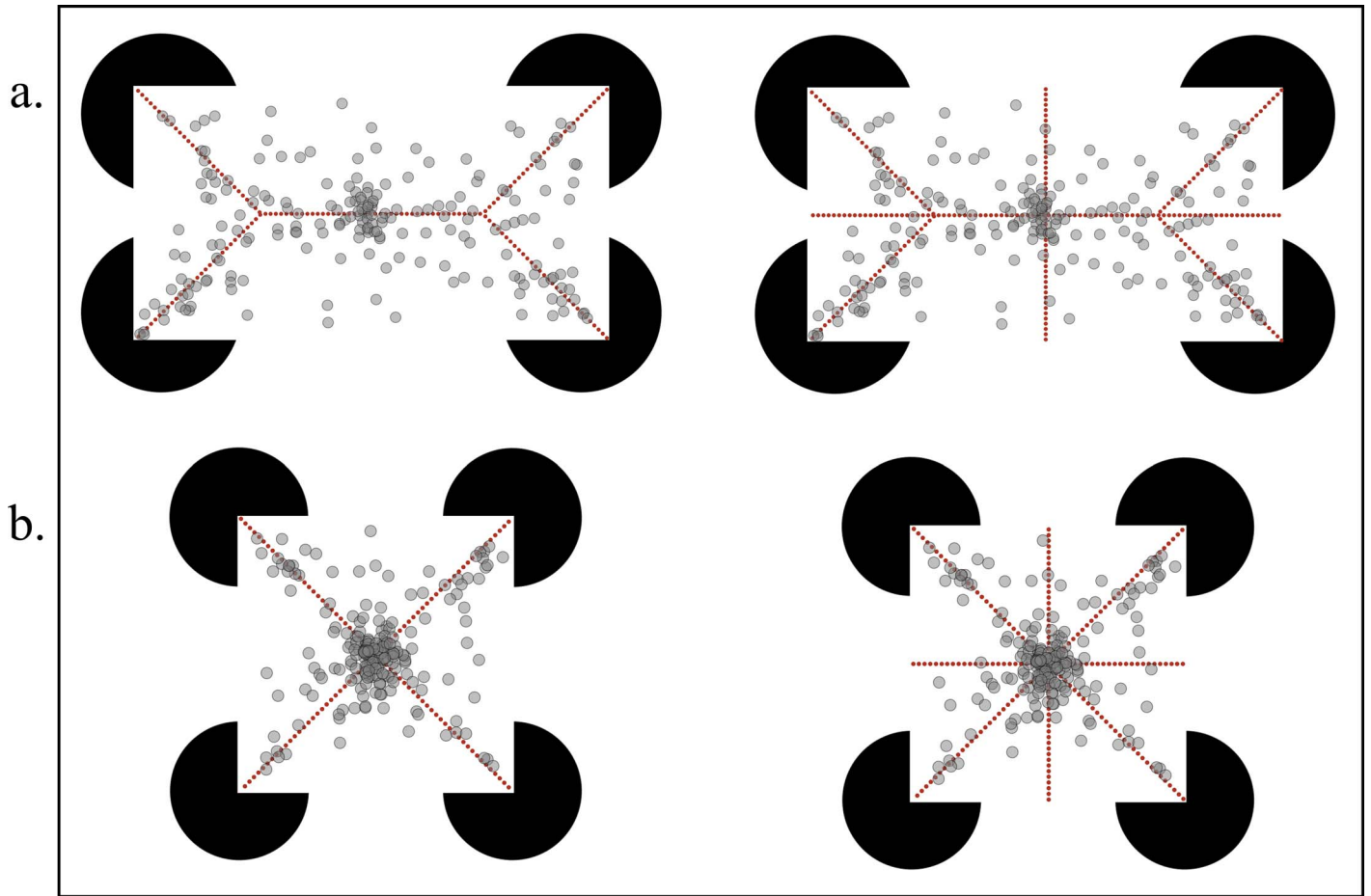
Figure 6. Conditions from Experiment 3: Kanizsa (a) rectangle and (b) square. Gray circles represent individual responses. Participants' responses are presented against pruned medial axes in these conditions (left column, red dashed lines) and the MAT computation (right column, red dashed lines). As described in the main text, stimuli in this experiment were presented against a white background on the tablet computer. Shapes are not drawn to scale.
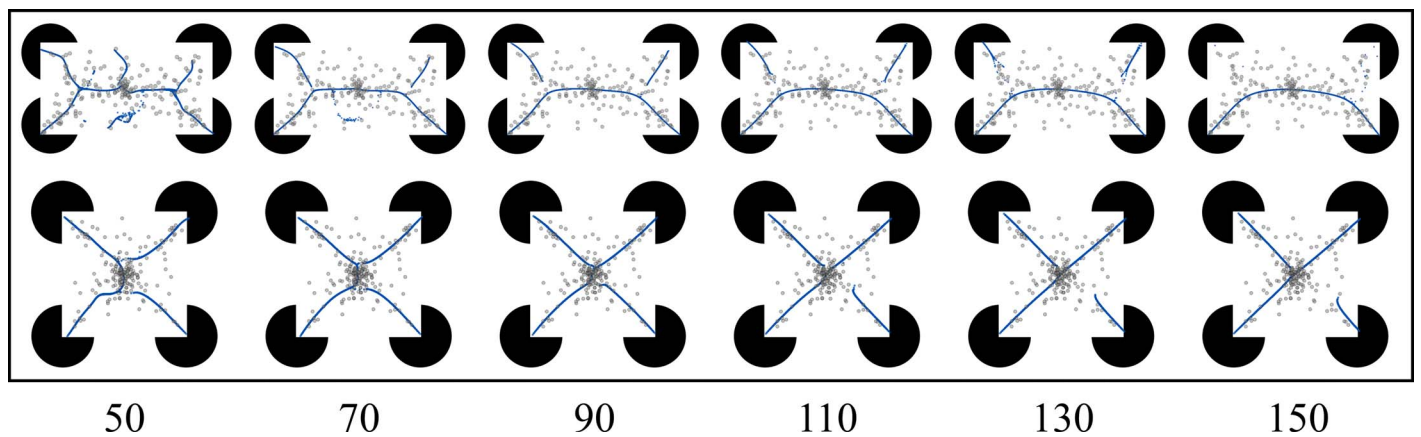

Figure 7. Smooth curves from the density ridge algorithm with increasing variance parameters. Each row displays a single shape from Experiment 3. Each column displays the smooth curves from each variance parameter (variance parameters displayed along the bottom). As described in the main text, stimuli in this experiment were presented against a white background on the tablet computer. Shapes are not drawn to scale.

Shape	Model	Comparison to chance		Model comparison			
		Response distance (SD)	Simulation distance	Log-likelihood	CI-low	CI-high	BIC
Rectangle	Pruned	40.88 (36.87)	66.40	−1582.85[a]	−1608.01	−1550.97	3176.30
	MAT	35.69 (34.11)	52.27	−1593.19[b]	−1617.90	−1562.76	3196.98
Square	Pruned	35.49 (32.69)	60.82	−1332.70[a]	−1367.57	−1293.03	2676.00
	MAT	34.29 (30.54)	49.25	−1340.41[b]	−1380.50	−1293.42	2691.41

Table 3. Results for Experiment 3. *Notes*: Mean participant response and simulation distances, as well as goodness-of-fit metrics, are displayed for each shape and axis structure. A mean response distance smaller than the simulation distance suggests that the model outperformed chance. Log-likelihood values indicate how well each axis structure fit participants' data, and confidence intervals (95% CI) provide estimates of uncertainty. For each axis structure, we display the results of the statistical model (center-controlled model: normal or t-distribution) that produced the smallest BIC value. [a] Log-likelihood and CIs using a center-controlled model following a t-distribution. [b] Log-likelihood and CIs using a center-controlled model following a normal distribution.

medial axis of Kanizsa shapes followed a pruned computation of uninterrupted contours, providing further support for medial axis pruning in humans. Although researchers have suggested that medial axis representations would incorporate Gestalt principles of perceptual completion (Johannes et al., 2001; Kimia, 2003), this hypothesis has not been empirically tested with human participants. As with perturbed shapes, participants' responses reflected a pruned medial axis, suggesting that skeletal representations describe an object's overall shape, rather than every contour. That participants' responses in the Kanizsa shapes reflected the medial axes of complete shapes further suggests that this task provides an assessment of shape perception, not simply a task-specific response strategy. Nevertheless, we acknowledge that the task instructions, which emphasized object shape (i.e., "tap the shape"), may have influenced participants' perception of the Kanizsa shapes and therefore their responses. Future research should consider minimizing task instructions further so as to compare medial axis models under conditions of illusory shape perception.

General discussion

Across multiple experiments, we found that participants' responses within 2-D shapes reflected the medial axes of the corresponding shapes. Participants' responses were better described by a medial axis structure than another viable model of shape perception or a task-specific cognitive strategy (Experiment 1). Moreover, the medial axis structure characterizing participants' responses reflected a lenient pruning computation under conditions of perturbation (Experiment 2) and illusory contours (Experiment 3). Our findings not only replicate previous work suggesting medial axis extraction using the tap-the-shape task (Firestone & Scholl, 2014; Psotka, 1978), but they also build upon this research by differentiating between

computations that accommodate every available edge and those that are more robust to noisy edges (Shaked & Bruckstein, 1998). By demonstrating how individuals extract the medial axis under conditions of perturbations and illusory contours, our findings help to bridge the gap between the many mathematical formulations of shape skeletons in computer vision (Wieser et al., 2017) and their potential biological implementation in human perception (Kimia, 2003).

Relation between the tap-the-shape task and visual shape perception

One might ask whether our conclusions about shape perception are justified, given that the task used to assess medial axis representation involved a tapping response, rather than a direct measure of shape perception. To address this concern, it was crucial to rule out alternative explanations for participants' responses. Here, we tested whether these results could be explained by a cognitive strategy specific to the tap-the-shape task, namely, a strategy to respond inside the white area and away from boundaries. As another test of visual shape representation, we also examined participants' responses in illusory shapes, and, finally, we included a statistical control for a known center bias in our model comparisons. In all cases, we found that the medial axis, particularly a leniently pruned model, provided the best fit to participants' responses.

Yet one additional concern is that participants may have used a response strategy that results in a pattern resembling the medial axis, but they did not extract the medial axis per se. For instance, instead of engaging in a boundary-avoidance strategy across the whole shape, participants' boundary-avoidance strategy could have been specific to local contour "neighborhoods" (e.g., between corners). This possibility makes predictions that are almost identical to the medial axis and is therefore difficult to rule out. However, one reason to believe that it does not account for our findings is that

such a strategy would more likely result in a response pattern similar to the MAT, not a pruned model, because the MAT describes all of the points in the shape equidistant from neighboring contours.

Nevertheless, it remains unclear why participants' taps would reflect the medial axes of different shapes in the first place. In other words, extraction of the medial axis by the perceptual system does not necessitate that participants' behavioral responses should reflect an underlying shape representation. One possible account for the observed behavior is that medial axis extraction leads to increased attention for locations along the medial axis of the shape. Thus, when participants are asked to simply tap a shape once, without other instructions, they may be biased to direct their taps to these perceptually enhanced locations. Single independent responses from different participants may capture the entirety of a medial axis structure by sampling across these locations. Similar perceptual enhancement has been observed for contrast sensitivity (Kovács et al., 1998; Kovács & Julesz, 1994) and texture segmentation (Harrison & Feldman, 2009) when stimuli are placed along the medial axis. Taken together, the results from these tasks suggest that medial axis extraction may occur automatically when participants are presented with a shape and that such extraction may influence behavior regardless of task relevance. Nevertheless, more research will be needed to fully understand the perceptual and cognitive mechanisms that influence participants' responses in the tap-the-shape task.

Determinants and implementation of medial axis pruning

The goal of pruning models is to provide a more accurate description of an object's shape by ignoring edges that may be considered noise. However, it remains unclear how the visual system determines which contours constitute noise. One determinant may be based on size. In the current study, we found evidence for a lenient pruning model that ignored both small and large perturbations but retained axial branches corresponding to the outer corners of the shape. On the one hand, our findings might suggest that size is irrelevant. On the other hand, relative to the corners of a shape, the perturbations used in the current study comprised a relatively small portion of the shape's geometry. Thus, the possibility remains that perturbations may affect the medial axis in direct proportion to the size of the local geometry, such that increasing the relative size of a perturbation would increase the likelihood that new branches emerge.

Another possible determinant for what constitutes as noise follows from rules governing object-part seg-mentation (Feldman et al., 2013). According to these rules, a perturbation may be less likely to be treated as noise if it is perceived as an object part. A particularly strong indicator of an object part is the presence of flanking points of concavity along the contour of a shape, separated by a short distance (De Winter & Wagemans, 2006; Hoffman & Richards, 1984; Singh, Seyranian, & Hoffman, 1999). Moreover, a candidate object part is less likely to be perceived as noise as it becomes larger in proportion to the rest of the shape (Dhandapani & Kimia, 2002). Although we did not test this hypothesis directly, participants' responses within the T-shape are consistent with it being a multipart shape, rather than a rectangle with extruding pertur-bations or, alternatively, a rectangle with two intruding perturbations. More specifically, the candidate parts of the T-shape are marked by flanking points of concav-ity. These points of concavity can be crossed by a relatively short distance. Finally, the part segmented by the points of concavity is comparable in size to the rest of the shape. By contrast, the rectangles with border perturbations in Experiment 2 had only single, non-flanking points of concavity by which to designate a part boundary (the inner corners of the perturbation), and the candidate part would be small relative to the rest of the shape. Thus, an intriguing possibility is that the degree to which medial axis branches grow or are pruned is related to whether perturbations delineate object parts and thus are less likely to be considered noise by the visual system (cf. Feldman et al., 2013).

Another determinant of the degree of pruning may depend on the perceptual detail required by the task. For instance, a skeletal structure consistent with the MAT might be elicited for subordinate-level categori-zation because it would describe the contours of a shape in more detail than a pruned computation. The type of lenient pruning observed in the current study may reflect a "default" medial axis wherein participants attend to the shape without engaging in a task that requires detailed shape discrimination.

The findings from the current study also raise questions about how pruning might be implemented by the human visual system. One possibility is that the visual system first computes a medial axis according to the MAT and then prunes away extraneous branches, with the result being the pruned skeleton observed in our data. This possibility is consistent with evidence of medial axis processing in early, edge-sensitive visual areas (e.g., Lee, 1996). According to this view, medial axis pruning occurs progressively as shape information is sent along the visual processing hierarchy but, crucially, reflects the structure of the MAT early in human vision. An alternative possibility is that the visual system extracts a medial axis consistent with pruning models from the outset of processing, in which case the term *pruning* may prove to be a misnomer. This

possibility is consistent with evidence of medial axis processing in higher-level visual areas that are less edge sensitive (e.g., Hung et al., 2012). According to this view, the medial axis is computed from a shape that has already been filtered for noisy contours or undergone perceptual completion (Kourtzi & Kanwisher, 2001) and thus has no need to extract additional branches. Some researchers have also suggested that a pruned medial axis may arise in early visual areas via top-down attentional mechanisms that extract only the relevant properties of the shape (Ardila et al., 2012). According to this account, feedback mechanisms may be recruited in certain contexts (e.g., subordinate-level categorization) to provide higher-resolution information about an object's contours (Lee, Mumford, Romero, & Lamme, 1998), such that, instead of being pruned, the medial axis "grows" new branches during visual processing to accommodate greater levels of detail. It is also possible that the aforementioned accounts are not mutually exclusive and that an object's shape may be represented simultaneously at multiple scales of detail by medial axes with various degrees of pruning (Green, 2017; Hummel, 2013).

Conclusion

Although it has long been known that shape perception is important for object recognition, it remains unclear how humans represent object shape, particularly in the presence of noisy contours. To the extent that the tap-the-shape task reveals the structure of human shape representations, we have provided evidence that humans extract a leniently pruned medial axis, a low-dimensional skeletal representation that is robust to various disruptions of shape. The results of the current study suggest that the medial axis, particularly a pruned formulation, may be important for creating a stable representation of an object's shape across contexts.

Keywords: shape skeleton, medial axis, visual perception, pruning, computation

Acknowledgments

All data, stimuli, and statistical models used in this study are available at https://osf.io/ac2pu/.

Corresponding authors: Vladislav Ayzenberg; Stella F. Lourenco.
Email: vayzenb@emory.edu; stella.lourenco@emory.edu.
Address: Emory University, Psychology, Atlanta, GA, USA.

Footnote

[1] A difficulty with comparing MAT and pruned models is that the added length of the MAT confers a statistical advantage when models are compared on the basis of response distance alone. To address this issue, Firestone and Scholl (2014) randomly sampled an equal number of points from each model. However, in comparisons of response distance, this approach does not account for the statistical advantage of the longer model. The points will still be distributed in a greater portion of the shape and will therefore be more likely to capture participants' responses by chance. For example, using this approach, a model consisting of evenly spaced points across the shape would outperform any of the axis models tested here because a participant's response would always be within close proximity of any point on the grid, regardless of where it fell. Therefore, as described in the main text, it is important that models be evaluated on the basis of their fit to participants' responses, such that the explanatory power of each axis point is considered in combination with response distance.

References

Ambosta, A. H., Reichert, J. F., & Kelly, D. M. (2013). Reorienting in virtual 3D environments: Do adult humans use principal axes, medial axes or local geometry? *PLoS One*, *8*(11), https://doi.org/10.1371/journal.pone.0078985.

Ardila, D., Mihalas, S., von der Heydt, R., & Niebur, E. (2012, March). *Medial axis generation in a model of perceptual organization.* Paper presented at the 2012 46th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ.

Attali, D., Boissonnat, J.-D., & Edelsbrunner, H. (2009). Stability and computation of medial axes-a state-of-the-art report. In T. Möller, B. Hamann, & R. D. Russell (Eds.), *Mathematical foundations of scientific visualization, computer graphics, and*

massive data exploration (pp. 109–125). Berlin: Springer.

Ayzenberg, V., & Lourenco, S. F. (2019). Skeletal descriptions of shape provide unique and privileged perceptual information for object recognition. *bioRxiv*, 1–22, https://doi.org/10.1101/518795.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147.

Blum, H. (1967). A transformation for extracting descriptors of shape. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 362–380). Cambridge, MA: MIT Press.

Blum, H. (1973). Biological shape and visual science (part I). *Journal of Theoretical Biology*, 38, 205–287.

Blum, H., & Nagel, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition*, 10, 167–180, https://doi.org/10.1016/0031-3203(78)90025-0.

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research*, 39, 257–269.

Bregman, A. S. (1981). Asking the "what for" question in auditory perception. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 99–118). Hillsdale, NJ: Erlbaum.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxberry.

Chaisilprungraung, T., German, J., & McCloskey, M. (2019). How are object shape axes defined? Evidence from mirror-image confusions. *Journal of Experimental Psychology: Human Perception and Performance*, 45, 111–124, https://doi.org/10.1037/xhp0000592.

Chen, Y.-C., Ho, S., Freeman, P. E., Genovese, C. R., & Wasserman, L. (2015). Cosmic web reconstruction through density ridges: Method and algorithm. *Monthly Notices of the Royal Astronomical Society*, 454, 1140–1156, https://doi.org/10.1093/mnras/stv1996.

Cheng, K., & Gallistel, C. R. (2005). Shape parameters explain data from spatial transformations: Comment on Pearce et al. (2004) and Tommasi & Polli (2004). *Journal of Experimental Psychology: Animal Behavior Processes*, 31, 254–259, https://doi.org/10.1037/0097-7403.31.2.254.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, UK: Cambridge University Press.

De Winter, J., & Wagemans, J. (2006). Segmentation of object outlines into parts: A large-scale integrative study. *Cognition*, 99, 275–325, https://doi.org/10.1016/j.cognition.2005.03.004.

Dhandapani, R., & Kimia, B. B. (2002, September). *Role of scale in partitioning shape*. Paper presented at the Proceedings of the International Conference on Image Processing, Rochester, NY.

Ebert, D., Brunet, P., & Navazo, I. (2002, May). *An augmented fast marching method for computing skeletons and centerlines*. Paper presented at the Proceedings of VisSym, Barcelona, Spain.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall.

Elder, J. H. (2018). Shape from contour: Computation and representation. *Annual Review of Vision Science*, 4, 423–450, https://doi.org/10.1146/annurev-vision-091517-034110.

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

Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., & Wilder, J. (2013). An integrated Bayesian approach to shape representation and perceptual organization. In S. J. Dickinson & Z. Pizlo (Eds.), *Shape perception in human and computer vision: An interdisciplinary perspective* (pp. 55–70). London: Springer.

Firestone, C., & Scholl, B. J. (2014). "Please tap the shape, anywhere you like" shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological Science*, 25, 377–386.

Giblin, P. J., & Kimia, B. B. (2003). On the local form and transitions of symmetry sets, medial axes, and shocks. *International Journal of Computer Vision*, 54, 143–157.

Green, E. J. (2017). A layered view of shape perception. *British Journal for the Philosophy of Science*, 68, 355–387, https://doi.org/10.1093/bjps/axv042.

Harrison, S. J., & Feldman, J. (2009). The influence of shape and skeletal axis structure on texture perception. *Journal of Vision*, 9(6):13, 1–21. https://doi.org/10.1167/9.6.13. [PubMed] [Article]

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, 18, 65–96, https://doi.org/10.1016/0010-0277(84)90022-2.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148, 574–591, https://doi.org/10.1113/jphysiol.1959.sp006308.

Hummel, J. E. (2013). Object recognition. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 32–46). New York: Oxford University Press.

Humphrey, G. K., & Jolicoeur, P. (1988). Visual object identification: Some effects of image foreshortening and monocular depth cues. In Z. W. Pylyshyn, *Computational processes in human vision: An interdisciplinary perspective* (pp. 430–444). Norwood, NJ: Ablex.

Humphrey, G. K., & Jolicoeur, P. (1993). An examination of the effects of axis foreshortening, monocular depth cues, and visual field on object identification. *Quarterly Journal of Experimental Psychology*, 46, 137–159.

Hung, C.-C., Carlson, E. T., & Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron*, 74, 1099–1113.

Huttenlocher, J., & Lourenco, S. F. (2007). Using spatial categories to reason about location. In J. M. Plumert & J. P. Spencer (Eds.), *The emerging spatial mind* (pp. 3–24). New York: Oxford University Press.

Johannes, M. S., Sebastian, T. B., Tek, H., & Kimia, B. B. (2001, July). *Perceptual organization as object recognition divided by two*. Paper presented at the Workshop on Perceptual Organization in Computer Vision, Vancouver, Canada.

Kanizsa, G. (1976). Subjective contours. *Scientific American*, 234, 48–52.

Kelly, D. M., & Durocher, S. (2011). Comparing geometric models for orientation: Medial vs. principal axes. *Communicative & Integrative Biology*, 4, 710–712.

Kimia, B. B. (2003). On the role of medial geometry in human vision. *Journal of Physiology-Paris*, 97, 155–190.

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, 189–224, https://doi.org/10.1007/bf01451741.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, 293, 1506–1509.

Kovács, I., Fehér, Á., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, 38, 2323–2333.

Kovács, I., & Julesz, B. (1994). Perceptual sensitivity maps within globally defined visual shapes. *Nature*, 370, 644–646.

Lee, T. S. (1996, July). *Neurophysiological evidence for image segmentation and medial axis computation in primate V1*. Paper presented at the Computation

and Neural Systems: Proceedings of the Fourth Annual Computational Neuroscience Conference, Location.

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology-Paris*, 97, 121–139.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, 38, 2429–2454.

Lescroart, M. D., & Biederman, I. (2013). Cortical representation of medial axis structure. *Cerebral Cortex*, 23, 629–637.

Leyton, M. (1989). Inferring causal history from shape. *Cognitive Science*, 13, 357–387, https://doi.org/10.1207/s15516709cog1303_2.

Leyton, M. (1992). *Symmetry, causality, mind*. Cambridge, MA: MIT press.

Li, Z. (2000). Can V1 mechanisms account for figure–ground and medial axis effects. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), *Advances in neural information processing systems* (Vol. 12, pp. 136–142). Cambridge, MA: MIT Press.

Liu, H., Wu, Z.-H., Zhang, X., & Hsu, D. F. (2013). A skeleton pruning algorithm based on information fusion. *Pattern Recognition Letters*, 34, 1138–1145, https://doi.org/10.1016/j.patrec.2013.03.013.

Liu, T.-L., & Geiger, D. (1999, September). *Approximate tree matching and shape similarity*. Paper presented at the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece.

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics*, 80, 1278–1289, https://doi.org/10.3758/s13414-017-1457-8.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B: Biological Sciences*, 200, 269–294.

Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Research*, 39, 2929–2946, https://doi.org/10.1016/S0042-6989(99)00029-2.

Montero, A. S., & Lang, J. (2012). Skeleton pruning by contour approximation and the integer medial axis transform. *Computers & Graphics*, 36, 477–487.

Ogniewicz, R. L., & Kübler, O. (1995). Hierarchic Voronoi skeletons. *Pattern Recognition*, 28, 343–359, https://doi.org/10.1016/0031-3203(94)00105-U.

Pizer, S. M., Oliver, W. R., & Bloomberg, S. H. (1987). Hierarchical shape description via the multiresolution symmetric axis transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

PAMI-9, 505–511, https://doi.org/10.1109/TPAMI. 1987.4767938.

Psotka, J. (1978). Perceptual processes that may create stick figures and balance. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 101–111.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464, https://doi. org/10.1214/aos/1176344136.

Sebastian, T. B., Klein, P. N., & Kimia, B. B. (2004). Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*, 550–571.

Shaked, D., & Bruckstein, A. M. (1998). Pruning medial axes. *Computer Vision and Image Understanding*, *69*, 156–169, https://doi.org/10.1006/cviu. 1997.0598.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., & Zucker, S. W. (1999). Shock graphs and shape matching. *International Journal of Computer Vision*, *35*, 13–32.

Singh, M., Seyranian, G. D., & Hoffman, D. D. (1999). Parsing silhouettes: The short-cut rule. *Perception & Psychophysics*, *61*, 636–660, https://doi.org/10. 3758/bf03205536.

Spröte, P., & Fleming, R. W. (2016). Bent out of shape: The visual inference of non-rigid shape transformations applied to objects. *Vision Research*, *126*, 330–346, https://doi.org/10.1016/j.visres.2015.08.009.

Spröte, P., Schmidt, F., & Fleming, R. W. (2016). Visual perception of shape altered by inferred causal history. *Scientific Reports*, *6*, 1–11, https:// doi.org/10.1038/srep36245.

Sturz, B., Boyer, T., Magnotti, J., & Bodily, K. (2017). Do eye movements during shape discrimination reveal an underlying geometric structure. *Animal Behavior and Cognition*, *4*, 267–285.

Telea, A., Sminchisescu, C., & Dickinson, S. (2004, September). *Optimal inference for hierarchical skeleton abstraction.* Paper presented at the Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK.

Trinh, N. H., & Kimia, B. B. (2011). Skeleton search: Category-specific object recognition and segmentation using a skeletal shape model. *International Journal of Computer Vision*, *94*, 215–240.

Vishwanath, D., & Kowler, E. (2003). Localization of shapes: Eye movements and perception compared. *Vision Research*, *43*, 1637–1653, https://doi.org/10. 1016/S0042-6989(03)00168-8.

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all-or-none samples. *Journal of Experimental Psychology: General*, *138*, 546–560, https://doi.org/10. 1037/a0017352.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, *19*, 645–647.

Wagemans, J., De Winter, J., de Beeck, H. O., Ploeger, A., Beckers, T., & Vanroose, P. (2008). Identification of everyday objects on the basis of silhouette and outline versions. *Perception*, *37*, 207–244.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*, 1172–1217.

Warrington, E. K., & James, M. (1986). Visual object recognition in patients with right-hemisphere lesions: Axes or features? *Perception*, *15*, 355–366.

Warrington, E. K., & Taylor, A. M. (1973). The contribution of the right parietal lobe to object recognition. *Cortex*, *9*, 152–164.

Wieser, E., Seidl, M., & Zeppelzauer, M. (2017). A study on skeletonization of complex petroglyph shapes. *Multimedia Tools and Applications*, *76*, 8285–8303, https://doi.org/10.1007/s11042-016-3395-1.

Appendix

Latent Variable Model 1: Basic Model With Equal Variance Assumed

This model assumes that participants' responses correspond to a point along a model's structure and some response (i.e., participants' taps) error. Response errors were assumed to be from a truncated normal distribution with equal variance around every point of the model's structure.

For each shape and structure, let S and T denote the two-dimensional subspace within a shape (i.e., feasible response region) and the axis points, respectively. This latent variable model assumes that each response is generated as follows:

STEP 1: Sample a point $\theta = (\theta_1, \theta_2)$ uniformly from T.
STEP 2: Given θ, sample a two-dimensional error $\varepsilon = (\varepsilon_1, \varepsilon_2)$, where ε_1 and ε_2 are independently sampled from a normal distribution with mean 0 and standard deviation σ.
STEP 3: Let $\mathbf{X} = \theta + \varepsilon$. If X is within the feasible region S, stop. Otherwise, repeat STEP 2, until $\mathbf{X}$ falls in S.

STEP 4: Output $\mathbf{X}$ as the response.

This model assumes that each participant's response is a slight deviation from the model's structure. More specifically, the point $\boldsymbol{\theta}$ in STEP 1 can be viewed as an ideal point on the model's structure that a participant ideally would like to touch. This ideal point is not observable, and therefore $\boldsymbol{\theta}$ is regarded as a latent vector. It is $\boldsymbol{\theta}$ plus some response error that is observable. Here, the error follows a truncated bivariate normal distribution that guarantees the observation $\mathbf{X}$ to be in the feasible region S. Note that σ^2 is the only unknown parameter in this model that quantifies the variation of the deviation from the axes. Based on this model, the marginal distribution of $\mathbf{X}$ has probability density function:

$$h(\mathbf{x}|\sigma^2) = \begin{cases} \frac{1}{|T|}\int_{\theta \in T} g(\mathrm{x}|\sigma^2;\boldsymbol{\theta})d\boldsymbol{\theta}, & \text{if } \mathbf{x} \in S, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbf{x} = (x_1, x_2)$, $|T|$ denotes the length of axis T and

$$g(\mathbf{x}|\sigma^2;\ \boldsymbol{\theta}) = \frac{\exp\left(-\frac{(x_1-\theta_1)^2+(x_2-\theta_2)^2}{2\sigma^2}\right)}{\int_{(y_1,y_2)\in S}\exp\left(-\frac{(y_1-\theta_1)^2+(y_2-\theta_2)^2}{2\sigma^2}\right)dy_1 dy_2}.$$

Under this latent variable model and given data $\mathbf{x}_1, ..., \mathbf{x}_n$ from n participants, the log-likelihood function of unknown parameter σ^2 is written as $l(\sigma^2) = \sum_{i=1}^n \log(h(\mathbf{x}_i|\sigma^2))$.

The maximum likelihood estimate of σ^2 is then obtained by $\hat{\sigma}^2 = \arg\max_{\sigma^2} l(\sigma^2)$, and the value of maximum likelihood function is $l(\hat{\sigma}^2)$. For each experiment, different structures were compared based on their maximum log-likelihood. The model with the largest maximum log-likelihood is regarded as the best-fitting model. To evaluate the uncertainty in $l(\hat{\sigma}^2)$, we also constructed a nonparametric bootstrap confidence interval (Davison & Hinkley, 1997; Efron & Tibshirani, 1994).

Latent Variable Model 2: Center-Controlled Model

We adopt the same notation as Latent Variable Model 1 (described above) and further denote $\mathbf{c} = (c_1, c_2)$ as the center of mass of the shape. The proposed model assumes that each response is generated as follows:

STEP 1: Sample a binary variable $D \in \{0, 1\}$, with $P(D = 1) = p$.

STEP 2: If $D = 1$, let $\boldsymbol{\theta} = \mathbf{c}$. Otherwise, sample a point $\boldsymbol{\theta} = (\theta_1, \theta_2)$ uniformly from T.

STEP 3: If $D = 1$, sample a two-dimensional error $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$, where ε_1 and ε_2 are independently sampled from normal distribution with mean 0 and standard deviation σ_1 or from a t-distribution with mean 0, scale parameter σ_1, and degree-of-freedom v. Otherwise, if $D = 0$, sample $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$, where ε_1 and ε_2 are independently sampled from normal distribution with mean 0 and standard deviation σ.

STEP 4: Let $\mathbf{X} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$. If $\mathbf{X}$ is within the feasible region S, stop. Otherwise, repeat STEP 3, until $\mathbf{X}$ falls in S.

STEP 5: Output $\mathbf{X}$ as the response.

This model extends Latent Variable Model 1 by allowing for additional weight at the center of the shape. According to this model, the participant either chooses the center of the shape (with probability p) or randomly chooses a point on a model's structure (with probability $1 - p$) as the ideal response point. Then, response error is generated depending on whether the ideal point is the center or sampled from the axis. We considered in the analysis two response error distributions, namely, a truncated normal distribution and a truncated t-distribution with one degree-of-freedom. The log-likelihood function for this model is a function of unknown parameters σ^2, σ_1^2, and p, written as

$$l(\sigma^2, \sigma_1^2, p) = \sum_{i=1}^n \log((1-p)h(\mathbf{x}_i|\sigma^2) + pf(\mathbf{x}_i|\sigma_1^2)),$$

where

$$f(\mathbf{x}|\sigma_1^2) = \frac{\exp\left(-\frac{(x_1-c_1)^2+(x_2-c_2)^2}{2\sigma_1^2}\right)}{\int_{(y_1,y_2)\in S}\exp\left(-\frac{(y_1-c_1)^2+(y_2-c_2)^2}{2\sigma_1^2}\right)dy_1\, dy_2}$$

when the tapping error distribution is assumed to be a truncated normal distribution and

$$f(\mathbf{x}|\sigma_1^2) = \frac{\left(\left(1+\frac{(x_1-c_1)^2}{v\sigma_1^2}\right)\left(1+\frac{(x_2-c_2)^2}{v\sigma_1^2}\right)\right)^{-\frac{v+1}{2}}}{\int_{(y_1,y_2)\in S}\left(\left(1+\frac{(y_1-c_1)^2}{v\sigma_1^2}\right)\left(1+\frac{(y_2-c_2)^2}{v\sigma_1^2}\right)\right)^{-\frac{v+1}{2}}dy_1\, dy_2}$$

when the tapping error distribution is assumed to be a truncated t-distribution with v degree-of-freedom ($v > 0$). Similar to the analysis of Model 1, parameter estimation and model comparison were conducted by making use of the log-likelihood function.