

Geometrically restricted image descriptors: A method to capture the appearance of shape.

Natalia Melnik, Daniel R Coates, Bilge Sayim

▶ **To cite this version:**

Natalia Melnik, Daniel R Coates, Bilge Sayim. Geometrically restricted image descriptors: A method to capture the appearance of shape.. Journal of Vision, 2021, 21 (3), pp.14. 10.1167/jov.21.3.14 . hal-03259331

HAL Id: hal-03259331
https://hal.univ-lille.fr/hal-03259331

Submitted on 14 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Geometrically restricted image descriptors: A method to capture the appearance of shape

Natalia Melnik Institute of Psychology, University of Bern, Bern, Switzerland

Daniel R. Coates Institute of Psychology, University of Bern, Bern, Switzerland and College of Optometry, University of Houston, Houston, Texas, USA

Bilge Sayim Institute of Psychology, University of Bern, Bern, Switzerland and Univ. Lille, CNRS, UMR 9193 – SCALab – Sciences Cognitives et Sciences Affectives, Lille, France

Shape perception varies depending on many factors. For example, presenting a stimulus in the periphery often yields a different appearance compared with its foveal presentation. However, how exactly shape appearance is altered under different conditions remains elusive. One reason for this is that studies typically measure identification performance, leaving details about target appearance unknown. The lack of appearance-based methods and general challenges to quantify appearance complicate the investigation of shape appearance. Here, we introduce Geometrically Restricted Image Descriptors (GRIDs), a method to investigate the appearance of shapes. Stimuli in the GRID paradigm are shapes consisting of distinct line elements placed on a grid by connecting grid nodes. Each line is treated as a discrete target. Observers are asked to capture target appearance by placing lines on a freely viewed response grid. We used GRIDs to investigate the appearance of letters and letter-like shapes. Targets were presented at 10° eccentricity in the right visual field. Gaze-contingent stimulus presentation was used to prevent eye movements to the target. The data were analyzed by quantifying the differences between targets and response in regard to overall accuracy, element discriminability, and several distinct error types. Our results show how shape appearance can be captured by GRIDs, and how a fine-grained analysis of stimulus parts provides quantifications of appearance typically not available in standard measures of performance. We propose that GRIDs are an effective tool to investigate the appearance of shapes.

Introduction

The appearance of a visual stimulus strongly varies under different viewing conditions. Even when a stimulus is correctly identified, knowledge about how it actually appeared to the observer remains limited because of the categorical nature of identification. Factors that influence how a stimulus appears are, for example, its presentation time (e.g., brief presentation: de Gardelle, Sackur, & Kouider, 2009; Fei-Fei, Iyer, Koch, & Perona, 2007; Johnson & Uhlarik, 1974; long presentation: Kanai, 2005; Troxler, 1804), its spatiotemporal context (e.g., masking: Sayim, Manassi, & Herzog, 2014; Suzuki & Cavanagh, 1998; serial dependence: Fischer & Whitney, 2014; Fritsche, Mostert, & de Lange, 2017), and its location in the visual field (Coates, Wagemans, & Sayim, 2017; Sayim & Wagemans, 2017). Here, we introduce a method that enables one to capture the appearance of shapes under various viewing conditions, allowing the quantification of appearance by considering stimulus parts as distinct targets. In particular, we introduce Geometrically Restricted Image Descriptors (GRIDs), and show by means of appearance variations in peripheral vision how GRIDs can be used to provide detailed characterizations of the appearance of shape.

Peripheral vision is distinct from foveal vision. For example, visual acuity in the periphery is reduced compared with the fovea (e.g., Anstis, 1998; Kerr, 1971; Mandelbaum & Sloan, 1947; Wertheim, 1894), and observers' abilities to detect blur (e.g., Wang & Ciuffreda, 2005), perceive color (e.g., Hansen, Pracejus, & Gegenfurtner, 2009; Parry, McKeefry, & Murray, 2006; Webster, Halen, Meyers, Winkler, & Werner, 2010), and detect image distortions (e.g., Bex, 2010) are compromised. Objects in the periphery also often appear ambiguous and indeterminate compared with foveal presentation (Baldwin, Burleigh, Pepperell, & Ruta, 2016; Bedell & Johnson, 1984; Coates et al., 2017;

Citation: Melnik, N., Coates, D. R., & Sayim, B. (2021). Geometrically restricted image descriptors: A method to capture the appearance of shape. *Journal of Vision*, *21*(3):14, 1–18, https://doi.org/10.1167/jov.21.3.14.

https://doi.org/10.1167/jov.21.3.14 Received June 24, 2020; published March 10, 2021 ISSN 1534-7362 Copyright 2021 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Downloaded from jov.arvojournals.org on 03/16/2021

Sayim, Myin, & Van Uytven, 2015; Sayim & Taylor, 2019; Sayim & Wagemans, 2017; Valsecchi, Koenderink, van Doorn, & Gegenfurtner, 2018; Valsecchi, Toscani, & Gegenfurtner, 2013; Yildirim, Coates, & Sayim, 2020). For example, the perceived size of peripherally presented targets was decreased and the shape was distorted compared with foveally presented targets (Baldwin et al., 2016; Newsome, 1972; Schneider, Ehrlich, Stein, Flaum, & Mangel, 1978; Thompson & Fowler, 1980). How exactly stimulus appearance differs between foveal and peripheral vision, however, remains unclear.

One of the key factors limiting peripheral vision is crowding, the interference of clutter (i.e., flankers) with target perception (Bouma, 1970, 1973; Herzog, Sayim, Manassi, & Chicherov, 2016; Levi, 2008; Manassi, Lonchampt, Clarke, & Herzog, 2016; Pelli & Tillman, 2008; Strasburger, Rentschler, & Juttner, 2011; Tripathy & Cavanagh, 2002; Whitney & Levi, 2011). Crowding hinders target identification (Coates, Bernard, & Chung, 2019; Greenwood, Sayim, & Cavanagh, 2014; Kooi, Toet, Tripathy, & Levi, 1994; Manassi, Sayim, & Herzog, 2013; Melnik, Coates, & Sayim, 2018, 2020; Pelli, Farell, & Moore, 2003; Rummens & Sayim, 2019; Saarela, Westheimer, & Herzog, 2010; Sayim & Wagemans, 2017) and alters target appearance (Coates et al., 2017; Greenwood, Bex, & Dakin, 2010; Sayim & Wagemans, 2017). A loss of target parts is often observed with crowded multisegment targets (target "diminishment"; Coates et al., 2017; Sayim & Wagemans, 2017). For example, Coates et al. (2017) presented the Rey-Osterrieth Complex Figure at different eccentricities (0°, 6°, and 12°), and asked participants to draw on a freely viewed response sheet what they perceived. Analyses of the drawings revealed that the target figure was captured less accurately the further away from fixation it was presented. In particular, the rate of target elements that were not depicted by observers (i.e., target diminishment) increased with increasing eccentricity. Similarly, in an appearance-based crowding paradigm investigating crowded letters, numbers, and letter-like targets, most errors owing to crowding were omissions of elements and truncations of target parts (Sayim & Wagemans, 2017). These errors were quantified by introducing error categories that captured how peripheral appearance diverged from accurate target perception: number (additions and omissions), length (extensions and truncations), position (translations and rotations), and shape (distortions), linking multidimensional variations of appearance with quantifiable performance measures.

How target appearance is related to identification performance is often unclear. For example, highly variable appearance of a given target across trials may only yield a single response in an identification task. Moreover, task demands and prior knowledge can strongly influence identification performance and conceal unbiased target appearance (e.g., Sayim & Taylor, 2019; see also Yildirim et al., 2020). For example, Sayim and Taylor (2019) presented observers with letter trigrams consisting of different or identical letters in the periphery. When the task was to identify the central letter of a three-letter string, target identification was highly accurate when three *T*s were presented. To investigate stimulus appearance, observers were asked to freely report and draw what they saw. Performance on the same stimulus (three *T*s) was poor: observers frequently missed one of the repeating items in the arrays of identical letters, verbally reporting and drawing only two letters ("redundancy masking"; see also Yildirim et al., 2020; 2021). Such mismatches between identification performance and captured appearance illustrates how key aspects of visual percepts can be missed in identification paradigms.

Here, we propose GRIDs, a method to investigate the appearance of shapes. We first introduce the basic features of the GRIDs method, and then show its application, and demonstrate a range of analyses that can be used on the collected data. GRIDs were used to design targets, and to collect responses. Target shapes were created by connecting nodes on a 3 × 3 square grid by line elements (Figure 1A). Observers were provided with printed response sheets displaying a grid (Figure 1B) and asked to capture target appearance by connecting points on the response sheet. The points on the response sheets constrain the possible locations for the placement of lines, limiting the degrees of freedom in comparison with, for example, free drawing paradigms (Barrett, Pacey, Bradley, Thibos, & Morrill, 2003; Coates et al., 2017; Hess, Campbell, & Greenhalgh, 1978; Johnson & Uhlarik, 1974; Metzger, 1936; Sayim & Wagemans, 2017), and enabling the treatment of each element as a target that can be reported correctly or incorrectly.

GRIDs can be applied in all contexts in which the appearance of shapes varies. Here, we used the GRIDs method to investigate the appearance of peripherally presented letters and letter-like targets. We quantified differences between the stimulus and response in three ways, analyzing overall accuracy and discriminability; the prevalence of number, length, and position errors; and the accuracy with which observers captured contour junctions. Moreover, we evaluated how accurately observers captured specific target features, such as horizontals, verticals, and obliques (see Appendix 1). Additionally, observers' fixation patterns were analyzed (see Appendix 2).

The captured appearance diverged from the presented targets. Letters were captured more accurately compared with letter-like targets, showing that performance depended on familiarity. GRIDs provide an advantage compared with usual performance measures, allowing to investigate the perception of each target element in detail. Each element in the

Downloaded from jov.arvojournals.org on 03/16/2021

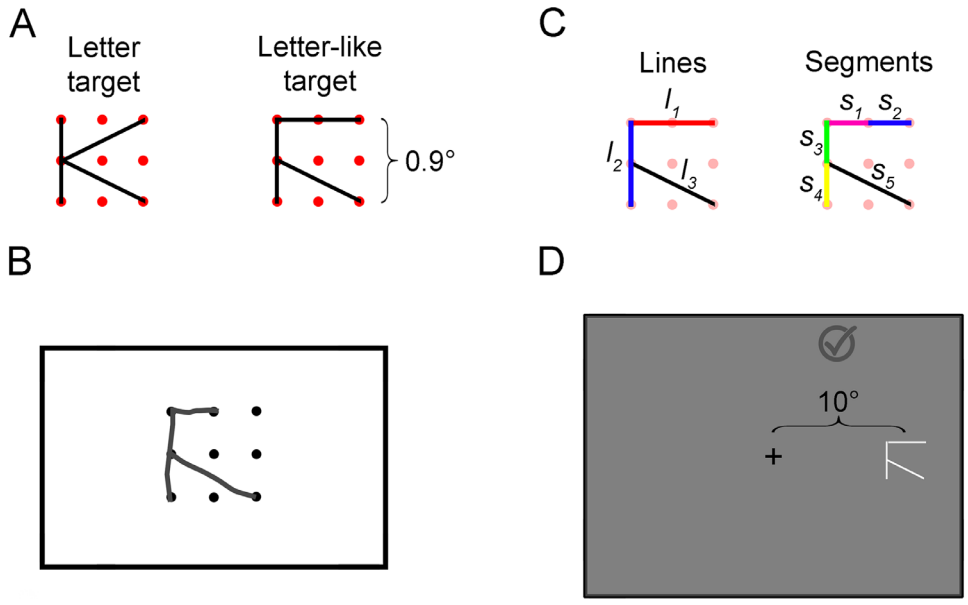


Figure 1. (A) Examples of a letter and a letter-like target as used in the experiment (the entire target set is shown in Table 1). The targets were created with lines and segments positioned on a 3 × 3 dot grid (shown in red for illustrative purposes; no dots were shown on the screen during the experiment). (B) An illustration of the dot grid used to record responses and a hypothetical response. (C) Examples of segments and lines (see Stimuli section for details). (D) Targets were presented at 10° in the right visual field when subjects fixated the central cross. When a trial was finished, observers fixated the checkmark symbol in the top part of the screen. (Display shown for illustrative purposes; the images are not to scale).

responses was characterized according to several error categories. Most errors occurred in letter-like targets, and only very few errors in letter targets. We found characteristic error patterns; however, strong variations existed between the different targets. Besides target familiarity, target shape was an important factor determining the number and types of errors. Similar to earlier studies that investigated complex, crowded stimuli (Coates et al., 2017; Sayim & Wagemans, 2017), we found high rates of truncations, indicating target diminishment with simple letter-like targets. The junction errors—again, mainly observed in letter-like targets—were characterized by a reduction of junction complexity: complex junctions (combinations of simpler junctions) were simplified and were rarely introduced as new junctions in the responses. Overall, our results indicate several distinctive categories of information loss in peripheral shapes that would be difficult to reveal with traditional forced-choice methods.

Methods

Participants

Ten observers participated in the experiment for course credit (1 male, 9 females; age range, 20–23 years; mean age, 21.2 years). All observers reported normal or corrected-to-normal visual acuity. Experiments were carried out with regard to ethical standards of the Declaration of Helsinki and were approved by the Ethics Committee of the University of Bern. Before the experiment, participants provided informed consent.

Apparatus

A 22-inch CRT monitor (HP p1230) set at a resolution of 1,152 × 864 pixels and a refresh rate of 110 Hz was used for stimulus presentation. Observers were supported with a chin and head rest placed at a distance of 57 cm from the screen. Eye movements were monitored using an Eye Link 1000 Plus eye tracker (SR Research Ltd., Mississauga, Ontario, Canada) at a 1000 Hz sampling rate. The eye tracker was positioned in a tower mount configuration (i.e., with the camera positioned above the working area). A combination of Python 2.7 and the PsychoPy toolbox (Peirce, 2007) was used for stimulus presentation and collection of behavioral and eye-tracking data. A7-sized paper (7.4 cm × 10.5 cm) with a dot grid printed in the middle was used to record observers' responses (Figure 1B). The response sheets were placed in front of the observer on an elevated board and were viewed from a distance of 46 to 50 cm (the size of the response sheets was about 8° × 12° of visual angle). The grid consisted of nine

Downloaded from jov.arvojournals.org on 03/16/2021

dots (diameter = 0.33 mm) arranged in a square 3 × 3 configuration with dots equally spaced at a distance of 0.75 cm (the overall extent of the grid on the paper was 1.5 × 1.5 cm, about 1.8°× 1.8° of visual angle). A standard pencil was used to record responses.

Stimuli

Stimuli consisted of 14 letter and 14 letter-like targets. Letters were selected from the Latin alphabet. Only letters that could be depicted on the 3 × 3 dot grid with straight lines were included (see Figure 1A; Table 1). The target set consisted of the letters A, E, F, H, K, L, M, N, T, V, W, X, Y, and Z. Each letter target was matched with one letter-like target, created by rearranging and/or rotating the elements of the corresponding letter (ten targets) or by inverting the entire letter (four targets). The numbers of elements (lines and segments; see next paragraph), perimetric complexity (perimeter over the ink area; Attneave & Arnoult, 1956), and the number of junctions in the two target sets were matched as closely as possible (Table 1).

The horizontal and vertical extent of the target grid was 0.9°. At least one line in each stimulus connected a leftmost and rightmost (or top and bottom) node of the grid (Figure 1A). We defined all parts of a target as segments and lines (Figure 1C). Segments were defined as strokes connecting two adjacent nodes without passing through any other dot (i.e., the smallest possible units on the grid), either horizontally or vertically (segment length = 0.45°) or diagonally (segment lengths = 0.64° and 1°). Lines connected two nodes on the grid, either horizontally/ vertically (line lengths = 0.45°, 0.9°) or diagonally (line lengths = 0.64°, 1°, 1.27°; Figure 1C), and contained one or two segments. Hence, all single segments that were not combined with a second segment (into a line) were also defined as lines (see Figure 1C, line l_1 contains two segments: s_1 and s_2, whereas line l_3 only contains segment s_5). Stimuli were white (72.07 cd/m^2), presented on a gray (34.5 cd/m^2) background. A black fixation cross (0.4° × 0.4°; 0.76 cd/m^2) was presented in the center of the screen.

Experimental task and procedure

The task was to replicate the appearance of the peripherally presented target as accurately as possible by connecting dots on the response grid. Because knowledge about the nature of the stimuli could influence how observers captured the appearance of the target, observers were not explicitly instructed about the nature of the stimuli. The target was presented at 10° eccentricity in the right visual field using a gaze-contingent presentation (Figure 1D). Targets were only displayed when the observers fixated a circular region with a radius of 2° around the fixation cross (the circular fixation region was not shown and not communicated to the observers). The observers were instructed not to look at the targets directly. They were allowed to view the target peripherally as long and often as necessary, looking back and forth between the response sheet and the fixation cross. To proceed to the next trial, observers fixated on a checkmark sign located at the upper right corner of the screen (Figure 1D). Observers could take breaks between the trials and were encouraged to take a break every 10 to 15 trials. Before each trial, a drift correction was performed. The eye tracker was calibrated at the beginning of the experiment and recalibrated as needed. No dots were shown on the screen during the experiment.

The order of the targets was randomized. To limit familiarization with the target set, each target was presented only once. Participants were familiarized with the procedure and performed several letter-like practice trials before starting the experiment. The targets used in the practice were not repeated in the experiment. Observers were not given feedback on their responses. Before the experiment, participants completed a similar task with crowded targets (not reported here). During the session, responses and eye tracking data were collected.

Analysis and results

We conducted two types of analyses. First, we evaluated the responses as a whole, quantifying overall accuracy and segment discriminability. Second, we analyzed the changes in the properties of lines (in terms of the length, number, and position of lines (see Corresponding Lines [CL] accuracy), and the changes in the junction types (see Junctions).

Overall accuracy and discriminability

Analysis

We evaluated the responses in terms of the overall accuracy. The responses were scored as correct if each segment of the target was replicated exactly and no segments were added. We also calculated the segment discriminability for each target—that is, how well the observers replicated each segment. To do so, we adapted the discriminability measure (d') from signal detection theory (Macmillan & Creelman, 2005). The correct placement of a segment was defined as a hit, and the absence of a segment was defined as a miss. To quantify false alarms and correct rejections, we computed the error distribution separately for each target. The error

Downloaded from jov.arvojournals.org on 03/16/2021

Target	Junctions							No. of Lines	No. of Segments	Symmetry	Perimetric Complexity	No. of Turns
	T	L	X	Y	K	F	V					
			2				1	3	4	+	12.39	11
			1				2	4	4		14.19	12
	1	2						4	8	+	16.99	9
	1	2						4	8	+	16.99	9
	1	1						3	6		12.98	7
	1	1						3	6		12.98	7
	2							3	6	+	12.74	8
	1		1					3	6	+	12.36	11
					1			3	4	+	14.78	7
	1	1						3	5		14.27	7
		1						2	4	+	9.24	4
							1	2	2		8.74	4
		1					2	4	6	+	12.81	8
	2	1						4	6	+	12.6	10
							2	3	6		12.86	6
		1				1		4	6		12.63	8
	1							2	4	+	8.99	5
	1							2	4	+	8.99	5
							1	2	2	+	11.74	4
							1	2	2	+	11.74	4
		1					2	4	6	+	12.81	8
		2					1	4	6		13.3	8
			1					2	4	+	9.95	8
			1					2	4		8.48	8
				1				3	3	+	7.64	6
							2	3	3		7.25	6
							2	3	6		12.86	6
			1				1	3	6		11.99	10
Overall	**0.36**	**0.43**	**0.21**	**0.07**	**0.07**	**0.00**	**0.71**	**2.93**	**4.93**	**0.78**	**12.06**	**76.93**
	0.50	**0.57**	**0.29**	**0.00**	**0.00**	**0.07**	**0.57**	**3.07**	**4.86**	**0.36**	**11.89**	**7.78**

Table 1. Characteristics of letters and letter-like targets. *Notes:* Numbers in the junctions, lines, and segments columns indicate the number of the corresponding features in the target. The + in the symmetry column indicates that the shape was symmetric. Perimetric complexity was calculated as the perimeter squared over an "ink" area (Attneave & Arnoult, 1956). The number of turns was calculated as the number of turns required to trace the outline of the figure (Attneave, 1957). We computed the number of turns by counting every instance of a change of direction in the outline (e.g., an angle or a termination point of a line).

Downloaded from jov.arvojournals.org on 03/16/2021

distribution contained all errors that actually occurred when the target was presented. The placement of a segment at any of the locations other than segment's true location resulted in a false alarm. A correct rejection was counted for each of the locations from the error distribution in which no segment was placed. (This method of calculating false alarms and correct rejections avoids the artificial increase of the number of "noise" segments that would result if all possible segment locations not occupied by the targets were considered as noise.) Discriminability was calculated using the equation $d' = z(H) - z(F)$, where $z(H)$ is the z-transformation of the hit rate and $z(F)$ is the z-transformation of the false alarm rate (Macmillan & Creelman, 2005). Bias (criterion) was computed using the formula $c = -1/2 * (z(H) + z(F))$. A negative bias indicated a tendency to leave the dots unconnected (i.e., not placing segments) and a positive bias indicated a bias toward placing more segments (i.e., connecting more dots). Because the z-transform reaches infinity when performance is 100% correct and 0% correct, values with 0% were replaced with 1% correct and values with 100% were replaced with 99% correct. Accuracy and segment discriminability provided simple measures of the degree to which the presented targets and segments were placed correctly.

We evaluated the extent to which the familiarity of the target (letter vs. letter-like) determined accuracy, segment discriminability, and bias with separate mixed-effect models. In all models, Familiarity of the target (letter vs. letter-like) were entered as a fixed factor, and subject and target identity (letter A, letter-like A, letter E, letter-like E, etc.) were entered as random factors. Because observers performed exceptionally well on the letters (see Results and Discussion), subsequent analyses focused on letter-like targets only. We also explored the influence of target complexity, measured by perimetric complexity (Attneave & Arnoult, 1956) and the number of turns (turns and changes in the outline of the shape as when tracing the figure; Attneave, 1957), on discriminability by computing Pearson's correlation coefficients. We manually quantified the number of turns in the responses by tracing the outline of the responses.

In addition to evaluating accuracy and discriminability as described above, we evaluated the similarity of the target and response images by pixel-wise correlations. With this analysis, strong (weak) correlations between two images indicate high (low) similarity of the images. We correlated the resulting correlation coefficients for each two images with segment discriminability ($r = 0.81$, $p < 0.001$) and accuracy ($r = 0.67$, $p < 0.001$). Both measures were strongly correlated with image similarity as measured by pixel-wise correlations, showing how pixel-wise correlations can be used as an additional, objective measure to quantify overall accuracy.

Results

Figure 2 shows the average discriminability (bars) and overall accuracy (line inserts) of each target. Average overall accuracy and discriminability were high ($d' = 3.88$, accuracy = 0.72). Inclusion of the target-type predictor improved the fit compared with a null model, accuracy: $\chi(1) = 12.68$, $p < 0.001$; d': $\chi(1) = 11.40$, $p < 0.001$. Responses were more accurate, and the discriminability was higher with the letters compared with letter-like targets, accuracy: 93.6% vs. 53%, $p < 0.001$; d': 4.46 vs. 3.29, $p < 0.01$. Overall, observers had low average biases in both target types (letter = 0.016, letter-like = –0.06; inclusion of the target-type predictor did not improve the fit compared with a null model, $p > 0.35$). However, the distribution of biases differed from target to target: biases were predominantly negative for the letter-like A, F, and Z targets, predominantly positive for the letter-like Y and N targets, and around zero for the remaining letter-like targets (Figure 2; Table 1).

We also compared global target characteristics (see Table 1 for lists of target characteristics, including global characteristics such as complexity and symmetry). In particular, assessing the relationship between perimetric complexity and discriminability of letter-like targets showed no correlation, $r(138) = 0.15$, $p = 0.08$; note that we did not parametrically vary target complexity (see also Appendix 3, Supplementary Figure 1). However, there was a negative correlation between the number of turns and discriminability of segments in letter-like targets, $r(138) = -0.30$, $p < 0.001$ (see also Appendix 3, Supplementary Figure 1), showing that discriminability increased as the number of turns of the target decreased. This correlation pattern was consistent across observers. We further explored the deviations between the number of turns in the letter-like targets and the responses. Our data showed an overall decrease of the number of turns in the letter-like targets (Appendix 3, Supplementary Figures 2 and 3). Taken together, these results showed profound differences between letters and letter-like targets, as well as between different letter-like targets.

CL accuracy

Analysis

The previous analyses focused on the differences between targets and responses by computing the accuracy and discriminability of the entire character irrespective of the correspondence between individual segments and lines in the responses to specific target segments and lines. Next, we investigated how accurately observers captured the properties of specific target segments and lines. Two trained raters (including the first author) assigned each line in the responses to a corresponding line in the target (we refer to this

Downloaded from jov.arvojournals.org on 03/16/2021

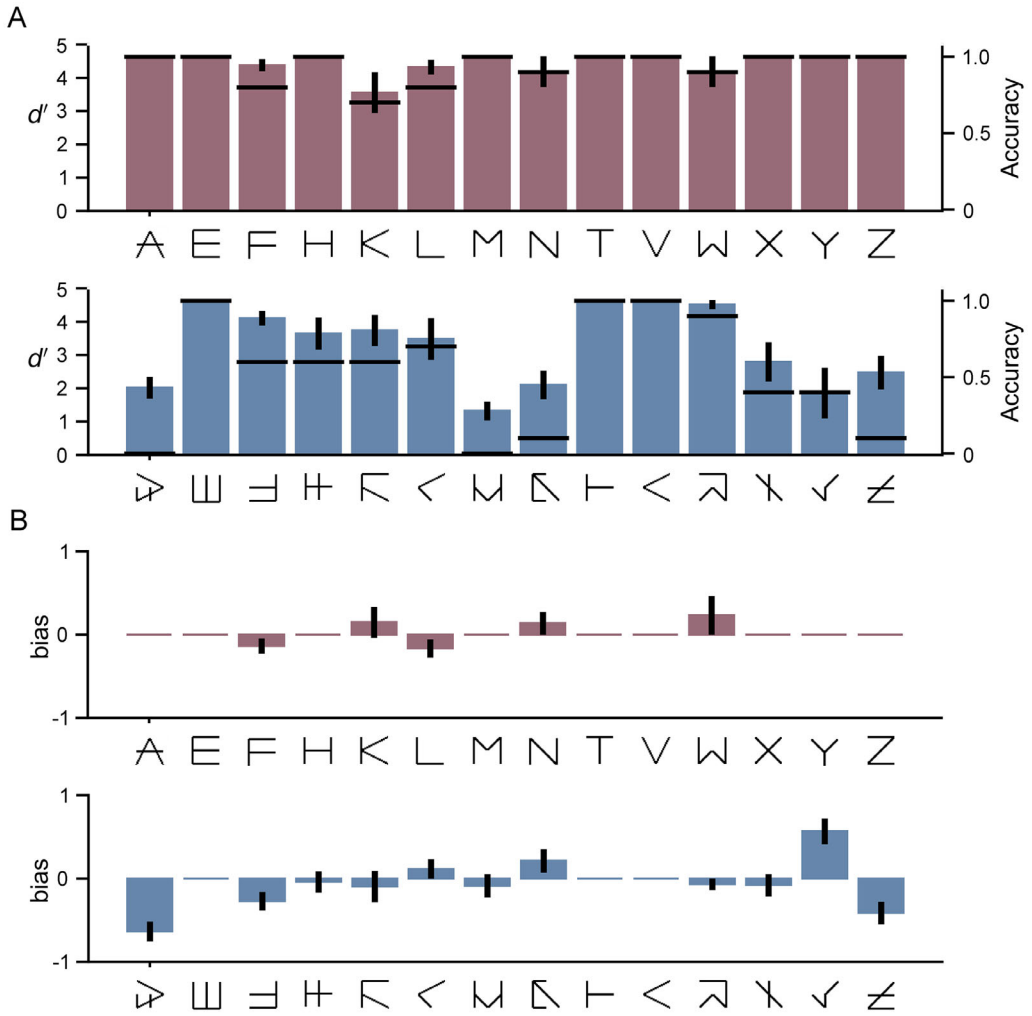


Figure 2. (A) Segment discriminability (d′; bars) and overall accuracy (black horizontal line inserts) for letter (familiar) and letter-like (unfamiliar) targets. (B) Bias for letter and letter-like targets. Negative bias denotes a bias to leave dots unconnected (i.e., not placing segments), positive bias denotes a tendency to place (more) segments. Error bars show standard error of the mean.

analysis with "CL" [corresponding lines] acronym). If a response line started and ended at the same points as the target line, the two lines were denoted as corresponding (Figure 3A; orange and cyan lines). The assignment of lines that did not match the target lines exactly could be performed using two hierarchical methods, prioritizing either 1) the orientation or 2) the location of the lines (see Figure 3A for an illustration of the approach). In both methods, the assignment was chosen that minimized the number of changes required to match the response to the target. Here, we used an orientation–primacy method for subsequent analyses (see Appendix 4 for the location–primacy method). Lines were denoted as corresponding when fulfilling the following criteria. First, the response line had to have a similar orientation as the target line. A similar orientation was defined as a tilt to the same side with less than 60° orientation difference (if the line was neither horizontal nor vertical), or a tilt to either side with less than 60° orientation difference (if the line in the target was horizontal or vertical). Second, the center of mass of the response line was shifted by one (e.g., translation of a line) or fewer (e.g., when the angle of a line changes) segment spaces (the minimal distances between two nodes on the grid) left, right, up, down, or diagonally.

The agreement between the two raters was calculated using Cohen's kappa (McHugh, 2012). Raters' agreement on line assignment was high (kappa = 0.98; the first author's ratings were used for the analysis). It took approximately 8 ± 2 seconds to categorize each response.

After the line assignment, we computed the deviations between the lines in the target and the assigned lines in the response to quantify CL accuracy. If the line in the response perfectly matched the line in the target (i.e., started and ended at the same points as the line in the target), no error was recorded. If the lines

Downloaded from jov.arvojournals.org on 03/16/2021

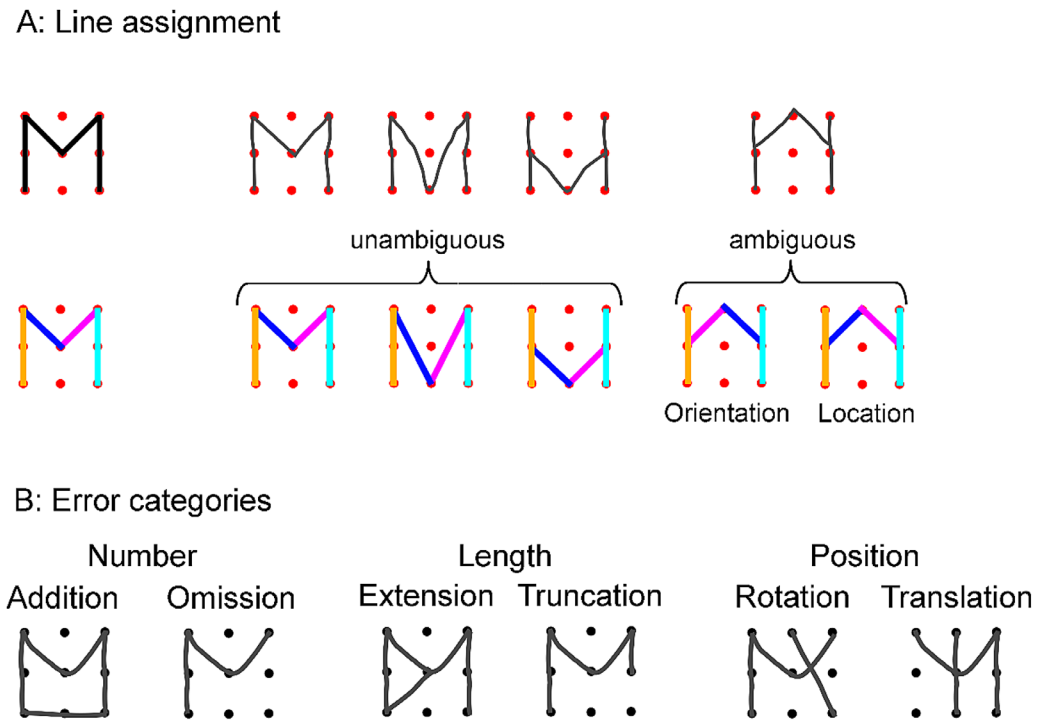


Figure 3. (A) Line assignment. Upper row: Illustrations of a target (M-letter) and exemplary potential responses. Lower row: Line assignment between the target and the responses. Corresponding lines are shown by the same color. Examples of unambiguous and ambiguous line assignments. In the ambiguous cases, prioritizing orientation and location is shown. (B) Error categories. Illustration of different error types for the M-letter target.

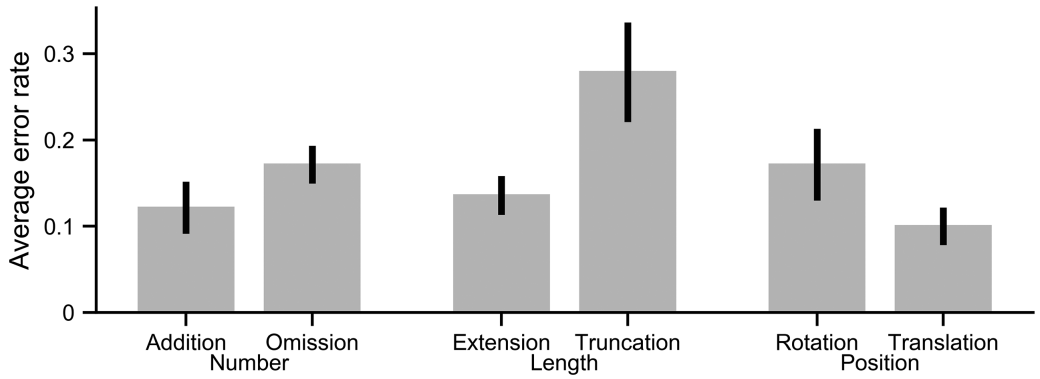


Figure 4. Average error rates of the letter-like targets for the number, length, and position error classes. Error bars denote standard errors of the mean.

were different than in the target, errors were recorded and labeled according to the following error categories (Sayim & Wagemans, 2017): number (addition [no corresponding line in the target] and omission), length (extension and truncation), and position (rotation and translation) (Figure 3B). When multiple errors occurred with the same line, all errors were scored (see, for example, a Rotation error combined with an Extension error in Figure 3B: when the line on the left is rotated, it is also extended).

We compared the average count of each error type with zero (i.e., no errors) to evaluate whether errors of that type occurred, using a one-sample Wilcoxon test. For all comparisons, p values were adjusted with a Bonferroni correction.

Results

Figure 4 shows the average error rates of each error type for the letter-like targets. Overall, subjects

Downloaded from jov.arvojournals.org on 03/16/2021

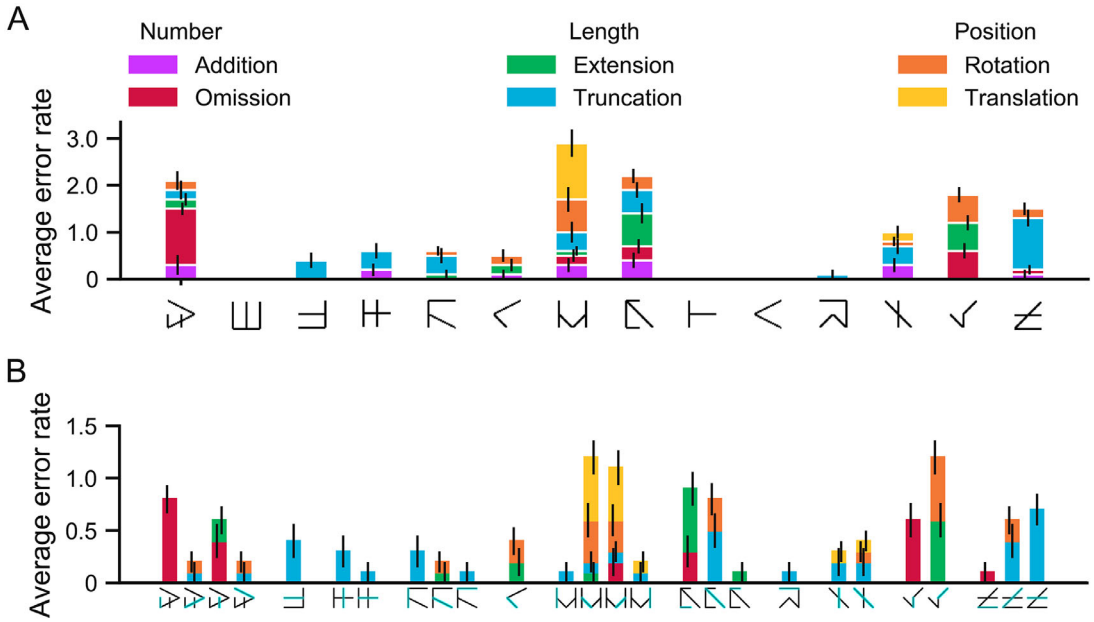


Figure 5 . Distribution of errors in the letter-like targets. (A) Errors averaged for each target. Absence of errors denote that all observers accurately captured that target. (B) Average errors shown separately for each target line. All lines that were not captured accurately are shown. Some errors were possible only as a combination of two errors (e.g., extension errors in the letter-like L target required concurrent rotation errors). Addition errors are not displayed in B as they do not directly correspond to any line. Error bars denote standard errors of the mean. (A small horizontal jitter was added to reduce the overlap of error bars.)

made errors of all error types (addition: $Z = 2.54$, $p = 0.047$, omission: $Z = 2.83$, $p = 0.012$; extension: $Z = 2.84$, $p = 0.012$; truncation: $Z = 2.67$, $p = 0.023$; rotation: $Z = 2.54$, $p = 0.047$; translation: $Z = 2.56$, $p = 0.047$). Figure 5 shows average error rates per target (Figure 5A) and per target line (Figure 5B). Errors were not homogeneously distributed over the targets (Figure 5A), ranging from no errors at all (letter-like E, T, and V) to all error types within a single target (letter-like M). Although truncation errors occurred for nearly all targets (except the letter-like Y and the perfectly reproduced targets), other errors, such as translations, were limited to only a small subset of targets and target lines (letter-like M and X targets). Most of the omission errors were observed in the letter-like A and Y targets, and most of the truncation errors in the letter-like Z target. Addition errors occurred in the letter-like A, H, M, N, and X targets. The errors were contingent on the properties of the lines in the target, including their length, orientation, and location (Figure 5B; see Appendix 5 for individual data). For instance, omission errors occurred frequently for short lines (in particular in the letter-like A and Y targets). Some of the observed omissions went along with spatial deformations. For instance, a vertical and a diagonal line in the letter-like Y target were often merged to form a (longer) diagonal line. Truncations occurred mainly for lines that were part of T- or X-junctions (e.g., letter-like F, H, K, N, and Z targets). Most of the extension errors were in the letter-like N and Y targets: In the letter-like N target, the majority of observers (six out of ten) extended the horizontal line to create a closed triangular shape, and in the letter-like Y target the majority of observers (six out of ten) extended the diagonal line while at the same time omitting the vertical line, indicating a merging of the two lines (see also Appendix 5). The inner, diagonal lines in the letter-like M target were often rotated or swapped (i.e., translation errors). Interestingly, we did not observe any rotation errors for vertical and horizontal lines. All rotation errors occurred with diagonal lines. By contrast, all other error types were also observed with vertical and horizontal lines.

Overall, the analysis of CL accuracy showed that captured appearance strongly diverged from the presented targets. The errors were not homogeneously distributed, showing the importance of the position, orientation, and length of individual lines within the target.

Junctions

Analysis

To quantify junction errors, we analyzed the differences between the junctions in the letter-like

Downloaded from jov.arvojournals.org on 03/16/2021

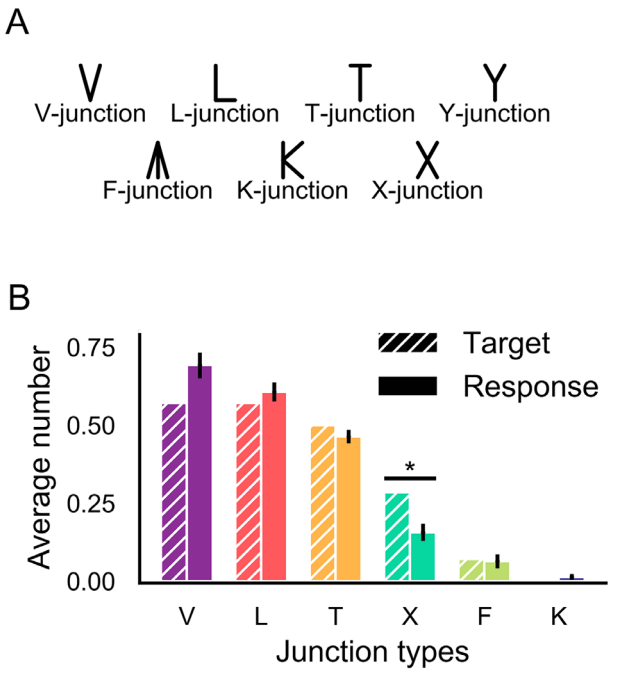


Figure 6. (A) Illustration of the junction types present in the letter-like target set. (B) The proportion of junctions in the targets (hashed bars) and responses (solid bars) for the six junction types. The error bars denote standard error of the mean. Significant differences between target and response junctions are indicated by asterisks (*$p < 0.05$).

targets and the responses. Five junction types that were in the target set (V-, L-, T-, F-, and X-junctions) and one junction type that was introduced in the response (K-junction) were included in the main analysis (see Figure 6A for examples of junction types and Table 1 for the distribution of the junctions across the targets). In addition to the main analysis, we also analyzed the data including only the L-, T-, and X-junctions. In this set, the V- and L-junctions were combined, K- and F-junctions were decomposed into two L-junctions.

The analyses used the raw line data (i.e., the overall number of junctions in the response) as well as the corresponding lines (see CL accuracy). First, we compared the average number of junctions in the targets with that in the responses using a one-sample Wilcoxon test (p values were adjusted using the Bonferroni procedure). Second, we evaluated whether junction changes in 'corresponding junctions' were different among the junction types. Corresponding junctions were defined as the junctions between the corresponding lines (see Analysis and Corresponding lines accuracy for details on how the corresponding lines were identified). We used three categories to classify the junction changes: additions, omissions, and transformations. Additions of junctions occurred when two or more lines formed a junction that was not present in the target; omissions occurred when a junction between corresponding target lines was not present in the response; and transformations occurred when a target junction had a different corresponding junction in the response.

Results

Figure 6B summarizes the average rate of each junction type in the target and in the response, disregarding their location. The most frequent junctions in our letter-like target set were V-, L-, and T-junctions (eight, eight, and seven junctions, respectively). There were four X-junctions and one F-junction. Because the junction types were not parametrically varied (i.e., the number of junctions of each junction type was not equal within the target set), we did not compare the results between junction types, but quantified differences between the targets and responses separately for each junction type. The junction distribution in the responses closely resembled the junction distribution in the targets. Overall, there were more simple junctions (V- and L-junctions) and fewer complex junctions (junctions that can be characterized as a combination of several junctions; e.g., T-junctions can be characterized as a combination of two L- or V-junctions). Observers accurately replicated the number of L-junctions ($Z = 1.44$, $p = 1$), T-junctions ($Z = -2.16$, $p > 0.18$), F-junctions ($Z = -0.33$, $p = 1$), K-junctions ($Z = 1.41$, $p = 1$), and V-junctions ($Z = 2.32$, $p > 0.11$). However, there was a difference between the number of junctions in the responses and the targets with X-junctions ($Z = -2.69$, $p < 0.023$). Observers' responses contained fewer X-junctions than were present in the targets (Figure 6B).

The second analysis quantified changes among corresponding junctions. The overall number of additions, omissions and transformation was similar (Figure 7A). V-, L-, and T-junctions were often added and V-, T-, and X-junctions were often omitted. Complex junctions (X-, F-, and T-junctions) were often transformed and simple junctions (V- and L-junctions) were rarely transformed (Figure 7B).

We also analyzed the data using only L-, T-, and X-junctions. The results were similar to those of the main analysis. Overall, we observed a similar trend for increase of L-junctions and decrease of X-junctions in the responses. Observers' responses contained fewer X-junctions than were present in the targets.

Overall, this exploratory analysis suggests a trend for a decrease of junction complexity in the periphery. Complex junctions were often simplified (e.g., X-junctions were transformed into T- and V-junctions). There was a strong relation between junction changes and the observed line/segment errors. Line omissions and truncations typically resulted in junction omissions and transformations to simpler junctions, and line

Downloaded from jov.arvojournals.org on 03/16/2021

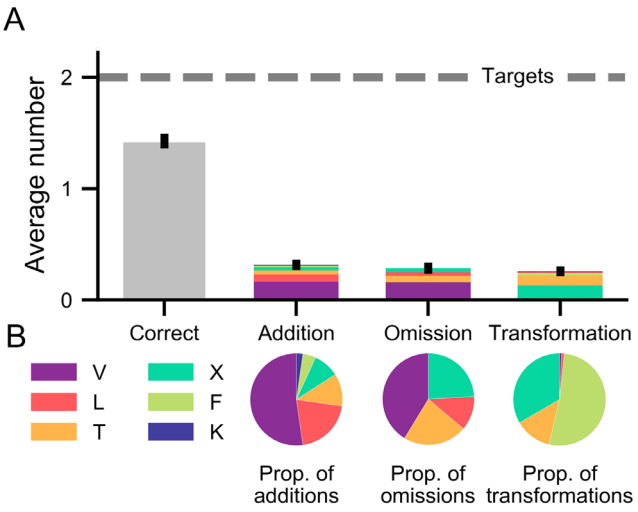


Figure 7. Junction changes (CL). (A) Average number of correct, added, omitted and transformed junctions in the responses. The dashed line indicates the average number of junctions in the targets. Color inserts on the bar graphs show the average (not normalized) proportions of added, omitted, and transformed junctions (proportion of additions: V: 52%, L: 20%, T: 11%, X: 9%, F: 5%, K: 2%; proportion of omissions: V: 55%, T: 20%, L: 12%, X: 12%; proportion of transformations: X: 55%, T: 33%, F: 11%, L: 3%, V: 3%). Note that the distribution of the junctions across the targets was not homogeneous (see Figure 6B and the text for details). (B) Proportions of added, omitted and transformed junctions. The proportions of the omitted and transformed junctions were normalized by the absolute number of those junctions in the target set. Overall, the majority of added and omitted junctions were simple junctions, and the majority of transformed junctions were complex junctions.

additions typically resulted in additions of junctions and transformations to more complex junctions.

Discussion

Appearance is a key product of shape processing by the human visual system. Thus, investigations of shape appearance can provide valuable insights of the mechanisms underlying shape perception. Besides the (usually unidimensional) method of adjustment, one approach to investigate stimulus appearance is to ask observers to verbally describe what they saw (Korte, 1923; Sayim & Taylor, 2019; see also Fei-Fei et al., 2007), another is to let observers draw how a target appeared (Barrett et al., 2003; Coates et al., 2017; Hess et al., 1978; Johnson & Uhlarik, 1974; Lettvin, 1976; Metzger, 1936; Sayim & Wagemans, 2017; Williams, 1985).

Using these methods, stimulus appearance has been characterized for a variety of target types, including basic shapes (Baldwin et al., 2016), gratings (Barrett et al., 2003; Hess et al., 1978), letters and letter-like characters (Sayim & Taylor, 2019; Sayim & Wagemans, 2017), and highly complex stimuli such as the Rey-Osterrieth complex figure (Coates et al., 2017). One key advantage of appearance-based methods (such as drawing and verbal description) compared with identification paradigms is that they make no (or minimal) assumptions in regard to the possible range of what an observer sees when being presented with a visual stimulus. For example, in contrast with standard letter identification, deviations from a certain letter or apparent morphs of letters can be recorded with GRIDs. At the same time, available methods to capture appearance have shortcomings. Free drawings and verbal descriptions can vary substantially between subjects. For instance, when targets are more complex, observers need to have a certain level of drawing proficiency to accurately depict how a target looked; for example, Coates et al. (2017) used art school students as participants, Sayim et al. (2015) a professional artist. The quantification of drawings often also requires additional steps such as setting thresholds to determine when an element is drawn correctly (see Sayim & Wagemans, 2017), in particular if no standardized scoring system is available (Coates et al., 2017).

Here, we introduced a method to capture the appearance of multisegment shapes that complements earlier appearance-based methods in several regards. First, GRIDs decrease the variation between subjects by reducing the degrees of freedom of line placements. Thus, subjects with different drawing skills can be included in a study, and the method can be used with any population that has basic motoric and visual skills. For example, using GRIDs with populations with disorders such as amblyopia (Levi, 2006) or dyslexia (Melnik, Coates, & Sayim, 2019) will be highly useful to capture how stimulus appearance differs in these populations. Second, in GRIDs, each stroke of the target shape can be treated as a discrete target. This allows a precise and detailed characterization and analysis of target parts, such as the accuracy of individual lines, line features, and junctions. Therefore, accuracy and, for example, confusions of different features can be compared directly without indirect measures of item confusions (e.g., Coates et al., 2019; see Appendix 3). Third, the method offers a large flexibility regarding the level of detail to be recorded. Different sizes and types of grids can be used to capture a variety of properties targeted by particular research questions. For example, for our exemplary task with letter and letter-like targets, using a 3×3 grid was sufficient to capture several key characteristics of target appearance, including basic differences between the letters and letter-like targets. The method can be

Downloaded from jov.arvojournals.org on 03/16/2021

adapted easily to study other simple or more complex shapes, such as typical outlines of shapes, provided that they consist of—or can be separated into—distinct parts. Fourth, with GRIDs (e.g., compared with free drawing), the responses can be analyzed directly based on grid coordinates and compared with the target without any additional processing, such as digitalization of the responses or setting the thresholds for what to consider a deviation from the target (Sayim & Wagemans, 2017; see also Baldwin et al., 2016; Barrett et al., 2003). GRIDs directly capture changes of all elements and their relations, including the sizes and proportions of shapes, yielding a clear advantage compared with free drawing paradigms. Although the resolution that can be captured with GRIDs is coarser than in typical free drawing paradigms and, therefore, may conceal smaller details of target appearance, its level of detail is greater than in a typical forced-choice paradigm. Importantly, the level of detail captured by the GRIDs is mainly constrained by observer's capacities in a given paradigm and not by any features of the method per se (i.e., any dot distance can be chosen).

In the present study, we showed the application of the GRIDs by quantifying the differences between targets and their appearance in peripheral vision captured with GRIDs. We analyzed the data using two types of analyses: by evaluating the targets as a whole, and by quantifying the changes and deviations in specific types of target properties (lines and junctions). We found a clear distinction in the perception of letters in comparison with letter-like targets: observers' performance was better with the letters than the letter-like targets. The captured appearance of letter-like targets strongly diverged from the presented targets. A variety of error types were observed in the responses, showing how appearance of shapes presented in the periphery diverges from free viewing. Similar to a study that investigated the appearance of crowded targets (Sayim & Wagemans, 2017), we observed truncations and omission of elements (target diminishment) among the most common errors. The observed diminishment was contingent on the position of the elements in the target (Figure 5). For example, when two lines in the target were positioned horizontally or vertically at the closest possible distance to each other (visual angle of 0.45°), observers perceived one of two lines as shorter than that in the target in 40% of the trials. A highly frequent truncation error was observed for the line that passes though the center of the grid in the letter-like Z target (in 70% of the trials). Omission errors often occurred for two (close by) short lines (corresponding to the shortest segments; see Figure 1C), with observers often omitting at least one of the two lines. For example, in the letter-like A target, observers missed at least one of the two vertical lines in 80% of the trials (60% one line, 20% both lines). The strong dependence of errors on the target type and spatial relations between lines, as well as the near absence of errors for simple targets (e.g., targets consisting of only two lines) suggest that the observed errors were due to crowding between the target parts ("self-crowding"; Martelli, Majaj, & Pelli, 2005; Zhang, Zhang, Xue, Liu, & Yu, 2009), or redundancy masking, the reduction of the perceived number of repeated elements (Sayim & Taylor, 2019; Taylor & Sayim, 2020; Yildirim et al., 2020; see also Taylor & Sayim, 2018), when the target lines were highly similar.

One explanation for the superior performance with letters is familiarity (Castet, Descamps, Denis-Noël, & Colé, 2017; Changizi & Shimojo, 2005; Krueger, 1975; Wiley, Wilson, & Rapp, 2016; Wong, Jobard, James, James, & Gauthier, 2009). Although we did not inform the observers about the nature of the stimuli, they could have formed a hypothesis that some targets were letters, potentially benefitting letter but not non-letter performance. However, observers rarely depicted any of the letters of the target set or distorted versions of them when shown letter-like stimuli (4% of responses to the letter-like trials), and never any other letter (as shown by visual inspection). Interestingly, although most omission errors occurred for identical lines, we did not find any evidence for redundancy masking (Sayim & Taylor, 2019; Taylor & Sayim, 2020; Yildirim et al., 2020; Yildirim, Coates, & Sayim, 2021) for the letter-like E target (i.e., reporting two instead of three horizontal lines). Because participants were presented with on average nine letters before the letter-like E, and the letter or letter-like F preceded the letter-like E for seven out of ten participants, expectation of a letter and implicit or explicit comparisons with the relatively easy F-target could underlie the absence of redundancy masking (see also Yildirim et al., 2021).

Even though letters and letter-like targets were matched as closely as possible in terms of number of segments, lines, junctions, and perimetric complexity, the two sets still differed in terms of other properties such as symmetry and types and frequency of junctions. For example, the letter set was more symmetrical compared with the letter-like target set: 78% of the letters, but only 36% of the letter-like targets were symmetrical around at least one axes of symmetry. Shape symmetry has been shown to enhance shape recognition (e.g., Carmody, Nodine, & Locher, 1977; Friedenberg & Bertamini, 2000; Kayaert & Wagemans, 2009; Machilsen, Pauwels, & Wagemans, 2009) and, thus, might be one of the reasons for increased accuracy for symmetrical (accuracy = 94.5% and 72.0% correct for letters and letter-like targets, respectively) compared with asymmetrical shapes (accuracy = 90.0% and 42.7% correct for letters and letter-like targets, respectively). In addition, the letter-like target set contained more targets made with short segments (segment length = 0.45° and 0.64°) that were not part of lines (six

Downloaded from jov.arvojournals.org on 03/16/2021

letter-like targets (letter-like A, L, M, N, W, and Y targets) versus three letter targets (letters M, W, and Y); Figure 1C). In the letter-like target set, errors were indeed less frequent for segments that were parts of lines compared with segments that were not part of lines (15% vs. 43% of segments, respectively).

Note that the difference between letter-like and letter targets was not due to a speed-accuracy trade-off. Analyses of the response completion durations and the peripheral target view durations (see Appendix 2) showed that letter-like trials took longer to be completed compared with the letter trials. Moreover, observers made more and longer peripheral views of the letter-like targets in comparison with the letter targets. Hence, differences in the response completion durations together with the higher error rates indicate a generally greater task difficulty with the letter-like targets compared with the letters.

Junctions have been proposed to be crucial for perception of shapes and scenes (e.g., Biederman, 1987; Corrow, Granrud, Mathison, & Yonas, 2012; Gibson, Lazareva, Gosselin, Schyns, & Wasserman, 2007; Rubin, 2001; Walther & Shen, 2014; Wilder, Dickinson, Jepson, & Walther, 2018). We evaluated how well the observers captured the junctions and whether junction changes were different among the junction types. Overall, the junction distributions in the responses resembled the junction distributions in the targets, reflecting the overall high levels of accuracy. Junction changes predominantly occurred in rather complex junctions, with a trend toward a decrease in complexity of those junctions in the responses. For example, the responses contained fewer X-junctions compared with the targets. An exploratory analysis of the changes among corresponding junctions showed that most of the omitted junctions were simple junctions, and most transformations of junctions occurred in complex junctions. A systematic variation of junctions is needed to further characterize the appearance of junctions of peripherally viewed shapes. Although we did not explicitly design our target set in the current study to investigate junctions (i.e., the number of junctions per each junction type was not equal), the GRIDs method is highly suitable to further investigate the appearance of junctions and their role in the perception of shapes.

Overall, the results of the present experiment showed several patterns that characterize shape perception in the periphery. With our stimuli, the types of errors that observers made were strongly linked to changes in complexity. The omission of one of the short lines of the letter-like target A, for example, resulted in a simplification of the shape in the response compared with the target (Figure 5; see also target diminishment; Coates et al., 2017; Sayim & Wagemans, 2017). Similarly, two lines with an obtuse angle between them (e.g., letter-like target Y in Table 1) were often combined into one line, again yielding a simplified version of the target. Responses were also simplified by creating closed shapes, for example, when an extended line was connected with another line in the response. The junction analysis showed that complex junctions were rarely added in the responses; however, when present in the target, they were frequently changed into simple junctions.

To assess this simplification, we compared the complexity of the targets and the responses measured by number of turns (the number of turns required to trace the outline of the figure; Attneave, 1957), and perimetric complexity (the perimeter squared over "ink" area; Appendix 3, Figures 2 and 3). A general trend was that letter-like targets with high numbers of turns (six or more) were affected more than letter-like targets with low numbers of turns. Overall, the number of turns in the responses tended to be lower than the number of turns in the targets (Appendix 3, Figure 3), suggesting a decrease in complexity in the responses compared with the targets. For example, the letter-like A was often missing one of the short vertical lines resulting in a decrease of the number of turns by two to four (depending on the missing line) compared with the target. By contrast, this type of simplification was rarely observed for simple targets with few turns. For example, in the letter-like T target, the same number of turns was always preserved in the responses. The average perimetric complexity followed a similar, albeit more varied, pattern. As expected, perimetric complexity was lower in the responses compared with the targets when omission and truncation errors were observed (Appendix 3, Figure 1). For example, although a shape became simpler (the number of turns decreased) when two lines merged—for example, for the letter-like Y target—perimetric complexity increased (the perimeter and the area of the shape changed). Interestingly, despite high perimetric complexity of the letter-like E target, all observers captured it well. This finding indicates that other aspects than perimetric complexity of the targets (e.g., symmetry and/or the Gestalt of the target) modulated perimetric complexity in the responses.

Two important characteristics of peripheral compared with foveal vision are its lower visual resolution and elevated crowding. Low visual resolution in the periphery (e.g., Anstis, 1998; Kerr, 1971; Mandelbaum & Sloan, 1947; Wertheim, 1894) could underlie some of the errors we observed. For example, the reported simplification is well in line with a reduction of access to high spatial frequencies (or blurring) of the stimuli when presented in the periphery. Several of the error categories and specific errors we reported would be expected by blurring a stimulus. In particular, omissions (especially of short lines) and truncations could be due to an apparent shortening of lines sufficient to either make the line go entirely

Downloaded from jov.arvojournals.org on 03/16/2021

undetected, or to become sufficiently unclear to not reach observers' criterion for including it in the response. Similarly, a simplification of junctions, as observed for many of the complex junctions, is expected when blurring a stimulus. However, blur does not easily account for other errors, such as extensions, translations and—the rather rare—additions. Orientation errors, by contrast, may result from the insufficient resolution of neighboring lines, for example, when two lines connected with a large angle are perceived as single straight line. Finally, blur alone does not predict the strong difference between letters and letter-like targets. However, it may be sufficient in many cases to increase uncertainty to a level that prevents a correct capture of unfamiliar—but not of familiar—shapes.

Crowding between the parts of our stimuli, or self-crowding (Martelli et al., 2005; Zhang et al., 2009), is equally a candidate that could have played a role in the observed errors. The strong accuracy variation between the different targets would, in this case, reflect their varying susceptibility to self-crowding. Instead of being caused by blur, the errors described above would be due to spatial interactions between the elements of a target. Owing to the strong configural differences and lack of systematic variation of the relations between the lines, definite conclusions about the role of self-crowding are difficult. However, several of the reported errors have been shown in crowding paradigms with more complex stimuli (Coates et al, 2017) and typical target–flanker configurations (Sayim & Wagemans, 2017), and may similarly be due to crowding in this experiment.

The observed position errors could be a consequence of faulty integration of detected features (e.g., Chung, Levi, & Legge, 2001; Greenwood, Bex, & Dakin, 2009; Levi, Hariharan, & Klein, 2002; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Pelli, Palomares, & Majaj, 2004), for instance, when features were correctly detected, but their positional information was compromised. This can be observed in the translation errors, for example, with the letter-like M target where the oblique lines in the target were swapped or changed their orientation (Figures 4 and 5). Finally, summary statistical representations which imply a loss of information similar to what we found here, may capture the observed appearance changes (e.g., Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011; Rosenholtz, Yu, & Keshvari, 2019). Because visualizations of the proposed perceptual effects owing to representations based on summary statistics can easily be generated for various kinds of stimuli, the resemblance of our responses with "mongrels" (synthesized images with identical summary statistics; Balas et al., 2009) can be tested. Future studies will shed light on the possible similarities and differences captured by appearance-based methods and representations based on summary statistics.

To conclude, we have introduced a new method to capture visual appearance. GRIDs are designed to capture the appearance of shapes at various levels of detail. Results can be analyzed qualitatively and by using performance measures, such as overall accuracy, segment discriminability, and line and junction accuracy. Differences between the presented targets and the responses captured with GRIDs indicate how a target appeared to an observer. Hence, the performance and performance differences reported here are a measure to quantify appearance. The versatility of the method allows its extension for usage in other paradigms in which target appearance is vague, indeterminate, and/or difficult (e.g., short presentation durations, masking, low contrast, and generally low visibility), and with various populations (e.g., elderly people, people with amblyopia, or people with dyslexia). Augmenting traditional measures for shape perception, the GRIDs method is an effective tool to investigate the appearance of shapes, and can help to shed light on how the visual system generates appearance from sensory inputs.

Keywords: shape perception, appearance, appearance-based methods, letter perception, peripheral vision, drawing, method of adjustment

Acknowledgments

The authors thank the two anonymous reviewers for their feedback.

Supported by the Swiss National Science Foundation (SNF; PP00P1_163723 awarded to Bilge Sayim).

The data from this study are available in the BORIS (Bern Open Repository and Information System) repository, https://boris.unibe.ch/id/eprint/144496. The experiment was not preregistered.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Parts of this work were presented at the 19th Annual Meeting of the Vision Sciences Society (2019) in St. Pete Beach, Florida, USA.

Commercial relationships: none.
Corresponding author: Natalia Melnik.
Email: natalia.melnik@psy.unibe.ch.
Address: Universität Bern, Institut für Psychologie, Fabrikstrasse 8, 3012 Bern, Switzerland.

Downloaded from jov.arvojournals.org on 03/16/2021

References

Anstis, S. (1998). Picturing peripheral acuity. *Perception, 27*(7), 817–825, doi:10.1068/p270817.

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology, 53*(4), 221–227, doi:10.1037/h0043921.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*(6), 452–471, doi:10.1037/h0044049.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12), 13–13, doi:10.1167/9.12.13.

Baldwin, J., Burleigh, A., Pepperell, R., & Ruta, N. (2016). The perceived size and shape of objects in peripheral vision. *I-Perception, 7*(4), 204166951666190, doi:10.1177/2041669516661900.

Barrett, B. T., Pacey, I. E., Bradley, A., Thibos, L. N., & Morrill, P. (2003). Nonveridical visual perception in human amblyopia. *Investigative Ophthalmology & Visual Science, 44*(4), 1555, doi:10.1167/iovs.02-0515.

Bedell, H. E., & Johnson, C. A. (1984). The perceived size of targets in the peripheral and central visual fields. *Ophthalmic and Physiological Optics, 4*(2), 123–131, doi:10.1111/j.1475-1313.1984.tb00345.x.

Bex, P. J. (2010). (In) sensitivity to spatial distortion in natural scenes. *Journal of Vision, 10*(2), 1–15, doi:10.1167/10.2.23.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115–147, doi:10.1037/0033-295X.94.2.115.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*(5241), 177–178, doi:10.1038/226177a0.

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research, 13*(4), 767–782, doi:10.1016/0042-6989(73)90041-2.

Carmody, D. P., Nodine, C. F., & Locher, P. J. (1977). Global detection of symmetry. *Perceptual and Motor Skills, 45*(3 Suppl), 1267–1273, doi:10.2466/pms.1977.45.3f.1267.

Castet, E., Descamps, M., Denis-Noël, A., & Colé, P. (2017). Letter and symbol identification: No evidence for letter-specific crowding mechanisms. *Journal of Vision, 17*(11), 2, doi:10.1167/17.11.2.

Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society B: Biological Sciences, 272*(1560), 267–275, doi:10.1098/rspb.2004.2942.

Chung, S. T. L., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research, 41*(14), 1833–1850, doi:10.1016/S0042-6989(01)00071-2.

Coates, D., Bernard, J.-B., & Chung, S. T. L. (2019). Feature contingencies when reading letter strings. *Vision Research, 156*, 84–95, doi:10.1016/j.visres.2019.01.005.

Coates, D., Wagemans, J., & Sayim, B. (2017). Diagnosing the periphery: Using the Rey–Osterrieth complex figure drawing test to characterize peripheral visual function. *I-Perception, 8*(3), 204166951770544, doi:10.1177/2041669517705447.

Corrow, S., Granrud, C. E., Mathison, J., & Yonas, A. (2012). Infants and adults use line junction information to perceive 3D shape. *Journal of Vision, 12*(1), 8–8, doi:10.1167/12.1.8.

de Gardelle, V., Sackur, J., & Kouider, S. (2009). Perceptual illusions in brief visual presentations. *Consciousness and Cognition, 18*(3), 569–577, doi:10.1016/j.concog.2009.03.002.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1), 10, doi:10.1167/7.1.10.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience, 17*(5), 738–743, doi:10.1038/nn.3689.

Freeman, J., & Simoncelli, E. P. (2011). *Metamers of the ventral stream*. *Nature Neuroscience, 14,* 1195–1201.

Friedenberg, J., & Bertamini, M. (2000). Contour symmetry detection: The influence of axis orientation and number of objects. *Acta Psychologica, 105*(1), 107–118, doi:10.1016/S0001-6918(00)00051-2.

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite effects of recent history on perception and decision. *Current Biology, 27*(4), 590–595, doi:10.1016/j.cub.2017.01.006.

Gibson, B. M., Lazareva, O. F., Gosselin, F., Schyns, P. G., & Wasserman, E. A. (2007). Nonaccidental properties underlie shape recognition in mammalian and nonmammalian vision. *Current Biology, 17*(4), 336–340, doi:10.1016/j.cub.2006.12.025.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding changes appearance. *Current Biology, 20*(6), 496–501, doi:10.1016/j.cub.2010.01.023.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences, USA, 106*(31), 13130–13135, doi:10.1073/pnas.0901352106.

Downloaded from jov.arvojournals.org on 03/16/2021

Greenwood, J. A., Sayim, B., & Cavanagh, P. (2014). Crowding is reduced by onset transients in the target object (but not in the flankers). *Journal of Vision, 14*(6), 2–2. doi:10.1167/14.6.2.

Hansen, T., Pracejus, L., & Gegenfurtner, K. R. (2009). Color perception in the intermediate periphery of the visual field. *Journal of Vision, 9*(4), 26–26, doi:10.1167/9.4.26.

Herzog, M. H., Sayim, B., Manassi, M., & Chicherov, V. (2016). What crowds in crowding? *Journal of Vision, 16*(11), 25, doi:10.1167/16.11.25.

Hess, R. F., Campbell, F. W., & Greenhalgh, T. (1978). On the nature of the neural abnormality in human amblyopia; neural aberrations and neural sensitivity loss. *Pflügers Archiv: European Journal of Physiology, 377*(3), 201–207, doi:10.1007/BF00584273.

Johnson, R. M., & Uhlarik, J. J. (1974). Fragmentation and identifiability of repeatedly presented brief visual stimuli. *Perception & Psychophysics, 15*(3), 533–538, doi:10.3758/BF03199298.

Kanai, R. (2005). Best illusion of the year contest: Healing grid. Retrieved from http://illusionoftheyear.com/2005/08/healing-grid/.

Kayaert, G., & Wagemans, J. (2009). Delayed shape matching benefits from simplicity and symmetry. *Vision Research, 49*(7), 708–717, doi:10.1016/j.visres.2009.01.002.

Kerr, J. L. (1971). Visual resolution in the periphery. *Perception & Psychophysics, 9*(3), 375–378, doi:10.3758/BF03212671.

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision, 8*(2), 255–279, doi:10.1163/156856894x00350.

Korte, W. (1923). Über die Gestaltauffassung im indirekten Sehen [On the apprehension of gestalt in indirect vision]. *Zeitschrift Für Psychologie, 93*, 17–82.

Krueger, L. E. (1975). Familiarity effects in visual information processing. *Psychological Bulletin, 82*(6), 949–974, doi:10.1037/0033-2909.82.6.949.

Lettvin, J. Y. (1976). On Seeing Sidelong. *The Sciences, 16*(4), 10–20, doi:10.1002/j.2326-1951.1976.tb01231.x.

Levi, D. (2006). Visual processing in amblyopia: Human studies. *Strabismus, 14*(1), 11–19, doi:10.1080/09273970500536243.

Levi, D. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Journal of Vision, 48*(5), 635–654, doi:10.1016/j.visres.2007.12.009..

Levi, D., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision, 2*(2), 3–3, doi:10.1167/2.2.3.

Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision, 9*(12), 11–11, doi:10.1167/9.12.11.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Manassi, M., Lonchampt, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision, 16*(3), 35, doi:10.1167/16.3.35.

Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision, 13*(13), 10–10, doi:10.1167/13.13.10.

Mandelbaum, J., & Sloan, L. L. (1947). Peripheral visual acuity with special reference to scotopic illumination. *American Journal of Ophthalmology, 30*(5), 581–588.

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5*(1), 6, doi:10.1167/5.1.6.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.

Melnik, N., Coates, D., & Sayim, B. (2018). Emergent features in the crowding zone: When target–flanker grouping surmounts crowding. *Journal of Vision, 18*(9), 19, doi:10.1167/18.9.19.

Melnik, N., Coates, D., & Sayim, B. (2019). What dyslexics see: Excessive information loss characterizes peripheral appearance in dyslexia. *Presented at the European Conference on Visual Perception, Leuven*.

Melnik, N., Coates, D., & Sayim, B. (2020). Emergent features break the rules of crowding. *Scientific Reports, 10*(1), 406, doi:10.1038/s41598-019-57277-y.

Metzger. (1936). *Gesetze des Sehens [Laws of seeing]*. Frankfurt, Germany: Waldemar Kramer.

Newsome, L. R. (1972). Visual angle and apparent size of objects in peripheral vision. *Perception & Psychophysics, 12*(3), 300–304, doi:10.3758/BF03207209.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–744, doi:10.1038/89532.

Downloaded from jov.arvojournals.org on 03/16/2021

Parry, N. R. A., McKeefry, D. J., & Murray, I. J. (2006). Variant and invariant color perception in the near peripheral retina. *Journal of the Optical Society of America A, 23*(7), 1586, doi:10.1364/JOSAA.23.001586.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13, doi:10.1016/j.jneumeth.2006.11.017.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12), 12, doi:10.1167/4.12.12.

Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature, 423*(6941), 752–756, doi:10.1038/nature01516.

Pelli, D. G, & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11*(10), 1129–1135, doi:10.1038/nn.2187.

Rosenholtz, R., Yu, D., & Keshvari, S. (2019). Challenges to pooling models of crowding: Implications for visual mechanisms. *Journal of Vision, 19*(7), 15, doi:10.1167/19.7.15.

Rubin, N. (2001). The role of junctions in surface completion and contour matching. *Perception, 30*(3), 339–366, doi:10.1068/p3173.

Rummens, K., & Sayim, B. (2019). Disrupting uniformity: Feature contrasts that reduce crowding interfere with peripheral word recognition. *Vision Research, 161*, 25–35, doi:10.1016/j.visres.2019.05.006.

Saarela, T. P., Westheimer, G., & Herzog, M. H. (2010). The effect of spacing regularity on visual crowding. *Journal of Vision, 10*(10), 17–17, doi:10.1167/10.10.17.

Sayim, B., Manassi, M., & Herzog, M. (2014). How color, regularity, and good Gestalt determine backward masking. *Journal of Vision, 14*(7), 8–8, doi:10.1167/14.7.8.

Sayim, B., Myin, E., & Van Uytven, T. (2015). Prior knowledge modulates peripheral color appearance. *Proceedings of the International Colour Association (AIC). Presented at the Tokyo Midterm meeting*, Tokyo, Japan, doi:10.7892/boris.85900.

Sayim, B., & Taylor, H. (2019). Letters Lost: Capturing Appearance in Crowded Peripheral Vision Reveals a New Kind of Masking. *Psychological Science, 30*(7), 1082–1086, doi:10.1177/0956797619847166.

Sayim, B., & Wagemans, J. (2017). Appearance changes and error characteristics in crowding revealed by drawings. *Journal of Vision, 17*(11), 8, doi:10.1167/17.11.8.

Schneider, B., Ehrlich, D. J., Stein, R., Flaum, M., & Mangel, S. (1978). Changes in the apparent lengths of lines as a function of degree of retinal eccentricity. *Perception, 7*(2), 215–223, doi:10.1068/p070215.

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision, 11*(5), 13–13, doi:10.1167/11.5.13.

Suzuki, S., & Cavanagh, P. (1998). A shape-contrast effect for briefly presented stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1315–1341, doi:10.1037/0096-1523.24.5.1315.

Taylor, H., & Sayim, B. (2018). Crowding, attention and consciousness: In support of the inference hypothesis. *Mind & Language, 33*(1), 17–33, doi:10.1111/mila.12169.

Taylor, H., & Sayim, B. (2020). Redundancy masking and the identity crowding debate. *Thought: A Journal of Philosophy, 9*(4), 257–265, doi:10.1002/tht3.469.

Thompson, J. G., & Fowler, K. A. (1980). The effects of retinal eccentricity and orientation on perceived length. *Journal of General Psychology, 103*(2), 227–232, doi:10.1080/00221309.1980.9921003.

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Research, 42*(20), 2357–2369, doi:10.1016/S0042-6989(02)00197-9.

Troxler, D. (1804). Über das Verschwinden gegebener Gegenstände innerhalb unseres Gesichtskreises [On the disappearance of given objects from our visual field]. *Ophthalmologische Bibliothek, 2*(2), 1–53.

Valsecchi, M., Koenderink, J., van Doorn, A., & Gegenfurtner, K. R. (2018). Prediction shapes peripheral appearance. *Journal of Vision, 18*(13), 21, doi:10.1167/18.13.21.

Valsecchi, M., Toscani, M., & Gegenfurtner, K. R. (2013). Perceived numerosity is reduced in peripheral vision. *Journal of Vision, 13*(13), 7–7, doi:10.1167/13.13.7.

Walther, D. B., & Shen, D. (2014). Nonaccidental properties underlie human categorization of complex natural scenes. *Psychological Science, 25*(4), 851–860, doi:10.1177/0956797613512662.

Wang, B., & Ciuffreda, K. (2005). Blur discrimination of the human eye in the near retinal periphery. *Optometry and Vision Science, 82*(1), 52–58, doi:10.1097/01.OPX.0000150185.69677.25.

Webster, M. A., Halen, K., Meyers, A. J., Winkler, P., & Werner, J. S. (2010). Colour appearance and compensation in the near periphery. *Proceedings of*

Downloaded from jov.arvojournals.org on 03/16/2021

the Royal Society B: Biological Sciences, 277(1689), 1817–1825, doi:10.1098/rspb.2009.1832.

Wertheim, T. (1894). Über die indirekte Sehschärfe [On indirect visual acuity]. *Zeitschrift Für Psycholologie & Physiologie Der Sinnesorgane, 7*, 172–187.

Whitney, D., & Levi, D. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*(4), 160–168, doi:10.1016/j.tics.2011.02.005.

Wilder, J., Dickinson, S., Jepson, A., & Walther, D. B. (2018). Spatial relationships between contours impact rapid scene classification. *Journal of Vision, 18*(8), 1, doi:10.1167/18.8.1.

Wiley, R. W., Wilson, C., & Rapp, B. (2016). The effects of alphabet and expertise on letter perception. *Journal of Experimental Psychology: Human Perception and Performance, 42*(8), 1186–1203, doi:10.1037/xhp0000213.

Williams, D. R. (1985). Aliasing in human foveal vision. *Vision Research, 25*(2), 195–205, doi:10.1016/0042-6989(85)90113-0.

Wong, A. C.-N., Jobard, G., James, K. H., James, T. W., & Gauthier, I. (2009). Expertise with characters in alphabetic and nonalphabetic writing systems engage overlapping occipito-temporal areas. *Cognitive Neuropsychology, 26*(1), 111–127, doi:10.1080/02643290802340972.

Yildirim, F., Coates, D., & Sayim, B. (2021). Hidden by bias: How standard psychophysical procedures conceal crucial aspects of peripheral visual appearance. *Scientific Reports, 11*(1), 4095.

Yildirim, F., Coates, D., & Sayim, B. (2020). Redundancy masking: The loss of repeated items in crowded peripheral vision. *Journal of Vision, 20*(4), 14, doi:10.1167/jov.20.4.14.

Zhang, J.-Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2009). Legibility of Chinese characters in peripheral vision and the top-down influences on crowding. *Vision Research, 49*(1), 44–53, doi:10.1016/j.visres.2008.09.021.

Downloaded from jov.arvojournals.org on 03/16/2021