

Appearance changes and error characteristics in crowding revealed by drawings

Brain and Cognition, University of Leuven (KU Leuven),
Leuven, Belgium
Institute of Psychology, University of Bern, Bern
Switzerland

SCALab (Sciences Cognitives et Sciences Affectives),
CNRS, University of Lille, Lille, France



Bilge Sayim

Brain and Cognition, University of Leuven (KU Leuven),
Leuven, Belgium



Johan Wagemans

Peripheral vision is strongly limited by crowding: Targets that are easily recognized in isolation are unrecognizable when flanked by close-by objects. Crowding does not only impair target recognition but also changes appearance. Here we investigated appearance changes and errors in crowding by letting observers draw crowded stimuli. Observers drew stimuli presented at 6° and 12° eccentricity. Stimuli consisted of characters and letter-like symbols. Targets were presented with either a flanker on each side or in isolation. To characterize appearance changes and errors in crowding, we developed a scoring system that captured differences between the drawings and the stimuli. The resulting drawings revealed strong appearance changes under crowding. Importantly, our results reveal crowding errors that are usually not shown in standard crowding paradigms. We found high rates of element Omissions and element Truncations, indicating a central role of target "diminishment" in crowding. Furthermore, we show that a subset of the observed element Omissions and Additions was possibly caused by feature migration. Relatively high rates of position errors, in particular element Translations, reflected the often reported location uncertainty in crowding. Virtually no complete target-flanker substitutions were observed. We suggest a new classification system for errors in crowding, and propose drawing as a useful appearance-based method to investigate crowding.

Bouma, 1976). Crowding is particularly strong in peripheral vision, and the spatial extent of crowding increases with eccentricity (e.g., Toet & Levi, 1992). Besides eccentricity, crowding is influenced by several factors, for example, target-flanker similarity (Kooi, Toet, Tripathy, & Levi, 1994; Sayim, Westheimer, & Herzog, 2008; Bernard & Chung, 2011; but see also Greenwood, Sayim, & Cavanagh, 2014), grouping (Banks & Prinzmetal, 1976; Livne & Sagi, 2007, 2010; Sayim, Westheimer, & Herzog, 2008, 2010, 2011; Saarela, Sayim, Westheimer, & Herzog, 2009; Manassi, Sayim, & Herzog, 2012, 2013; Sayim & Cavanagh, 2013; Herzog, Sayim, Chicherov, & Manassi, 2015), and flanker position (e.g., inward-outward anisotropy: Bouma, 1970; Banks, Bachrach, & Larson, 1977; Petrov & Popple, 2007; radial-tangential asymmetry: Toet & Levi, 1992; Pelli et al., 2007).

There are several, not necessarily mutually exclusive explanations of crowding which try to accommodate this rich scope of findings. For instance, spatial pooling and averaging (Wilkinson, Wilson, & Ellemberg, 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), excessive feature integration (Pelli, Palomares, & Majaj, 2004), limits of attentional resolution (He et al., 1996; Intriligator & Cavanagh, 2001), and imprecise attentional selection (Chastain, 1982; Strasburger, Harvey, & Rentschler, 1991) have been proposed to underlie crowding. A unified account of the processes underlying crowding, however, is still lacking (e.g., Levi, 2008; Whitney & Levi, 2011).

One way to evaluate different accounts of crowding is to investigate crowding errors (e.g., Huckauf & Heller, 2002; Freeman, Chakravarthi, & Pelli, 2012; Hanus & Vul, 2013). In particular, reduced performance in crowding may be due to erroneous combi-

Introduction

In crowding, a target is more difficult to identify when it is flanked by close-by items compared to when presented in isolation (e.g., Bouma, 1970; Andriessen &

Citation: Sayim, B., & Wagemans, J. (2017). Appearance changes and error characteristics in crowding revealed by drawings. *Journal of Vision*, 17(11):8, 1–16, doi:10.1167/17.11.8.

doi: 10.1167/17.11.8

Received February 18, 2015; published September 28, 2017

ISSN 1534-7362 Copyright 2017 The Authors



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

Downloaded From: <http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/936469/> on 10/10/2017

Error class	Number		Size		Position		Shape	Substitution
Error type	Omission	Addition	Truncation	Extension	Translation	Rotation	Distortion	

Table 1. Error classes and error types. Drawings were evaluated in regard to the number, size, position (linear and angular), and shape of elements (i.e., element errors), as well as target-flanker substitutions.

nations of features, target-flanker substitution, target suppression, or random errors. For example, it was shown that observers asked to report the orientation of a Gabor target were inclined to report the average orientation of the target and the flankers (Parkes et al., 2001), indicating that individual target information was unavailable. Instead, target and flanker features were erroneously combined (see also Pöder & Wagemans, 2007). Substitution errors occur when an entire flanker (Strasburger et al., 1991) or features of a flanker (e.g., Wolford, 1975; Wolford & Shum, 1980) are substituted for a target or target features. For example, observers frequently report a flanker instead of the target (e.g., Chung & Legge, 2009; Strasburger et al., 1991; Strasburger, 2005). Reduced performance in crowded compared to uncrowded conditions may also be due to target suppression by the flankers, i.e., a reduction of target visibility (e.g., Estes, 1972). Finally, “random errors,” for example lapses, are errors that are not specific to crowding.

Crowding not only deteriorates performance but also changes appearance (Greenwood, Bex, & Dakin, 2010; Coates, Wagemans, & Sayim, 2017), and phenomenological reports vividly describe the effects of crowding on appearance (e.g., Korte, 1923). How to best capture the appearance of crowded stimuli in a more quantitative fashion, however, is an open question. One way to approach the influences of crowding on appearance is to investigate stimuli that are physically different but appear the same (see e.g., Koenderink, Valsecchi, van Doorn, Wagemans, & Gegenfurtner, 2017). For example, it was proposed that the visual system represents peripheral images by extracting a set of summary statistics from the visual periphery (Balas et al., 2009; see also Freeman & Simoncelli, 2011). As the visual system only has access to the summary statistics, stimuli with the same statistics should be indiscernible, i.e., appear the same. Hence, crowding errors would manifest as the failure to distinguish physically different stimuli with the same summary statistics.

Detailed accounts of appearance changes in crowding, however, are still lacking. In particular, it remains unclear how the appearance of complex stimuli, such as letters, is influenced by crowding, and whether and to what degree such appearance changes are systematic. Previous coarse classifications of possible error types (feature combination, substitution, etc.) are not suited to capture more specific appearance changes in crowding. Therefore, in order to give a more detailed

account of crowding errors, we developed a drawing method which captures appearance changes and errors in crowding. Participants drew crowded stimuli presented in the visual periphery. We used drawing as our method of choice to avoid restricting responses to predefined response categories. The drawing method permits to capture details of appearance changes in crowding that often go unnoticed when using performance measures with a limited number of response categories.

Stimuli consisted of characters and letter-like symbols, presented at 6° and 12° eccentricity. Eye tracking assured that stimuli were only presented when participants fixated a central fixation dot. In Experiment 1, participants were presented with letters, numbers, and letter-like symbols, flanked by two flanker types of different complexity. In Experiment 2, all targets and flankers differed from each other, and each item consisted of a unique, meaningless configuration of lines. We developed a scoring system to evaluate the resulting drawings. The drawings were analyzed by quantifying the differences between the presented stimuli and the drawings. We used five error classes with eight different error types (Table 1) to capture the differences between presented and depicted stimuli. Taking the drawings as representations of peripheral appearance, we found evidence for strong changes of appearance when the stimuli were crowded. The most common errors were “diminishment” errors (whereby elements were often missing or truncated in the drawings), and position errors (especially translations of elements when crowding was strong). Substitutions of entire targets/flankers were virtually absent in both experiments, showing that such complete substitutions did not play a role in our experiments. However, we found evidence for element migration between the target and the flankers. In particular, additional elements drawn at the target location were frequently accompanied by element Omissions in the flankers (especially in Experiment 2). This suggests that the additional elements were not perceptually generated but perceived at the wrong location. By characterizing and quantifying errors in crowding, we show how crowding changes appearance. We propose that the drawing method is useful to investigate crowding as it may reveal details of appearance changes and errors in crowding that go unnoticed in standard crowding paradigms.

Experiment 1

Materials and methods

Observers

Five observers participated in the experiment. One additional observer was excluded because of an erroneous order of the drawings. All observers were art students recruited from a local art school. Participants were naive as to the purpose of the experiment. All observers reported normal or corrected-to-normal visual acuity. The experiments were carried out according to ethical standards specified in the Declaration of Helsinki and were approved by the Ethics Committee of the KU Leuven. Before the experiment, participants gave informed consent.

Apparatus

Stimuli were presented on a Sony Trinitron GDM-F520 CRT monitor driven by a standard accelerated graphics card. The screen resolution of the CRT was set to 1152 by 864 pixels. The refresh rate was set to 120 Hz. Observers were supported by a chin and head rest and viewed the monitor from a distance of 57 cm. Eye movements were recorded using an SR Research EyeLink 1000 at 1000 Hz sampling rate. An elevated drawing board was placed in front of the head and chin rest. A drawing book was placed on the board, and an electronic pen was used for drawing. This setup enabled viewing the screen and the drawing book without head movements, and without blocking the camera of the eye tracker. An electronic pen has the advantage that it yields analogue and digital results at the same time, and does not require special training as, for example, drawing on a tablet computer does. The experiment was programmed in MATLAB (Mathworks, Natick MA) in combination with the Psychophysics toolbox (Brainard, 1997).

Stimuli

Stimuli consisted of characters and letter-like symbols (Figure 1). Targets consisted of a range of numbers, letters, symbols, and rotated numbers and letters, each consisting of three basic line elements (either three lines, or an arc and two lines, used to generate the targets F, K, R, Z, 2, 4, and 5). Targets were presented alone (*baseline* condition) or flanked on each side by a flanker. There were two types of flankers. First, flankers consisted of a single vertical bar (*simple flanker* condition). Second, two vertical bars crossed the horizontal bar making up a double-dagger symbol (*complex flanker* condition). The line elements were 0.05° wide. All items had a vertical extent of 1.2° and

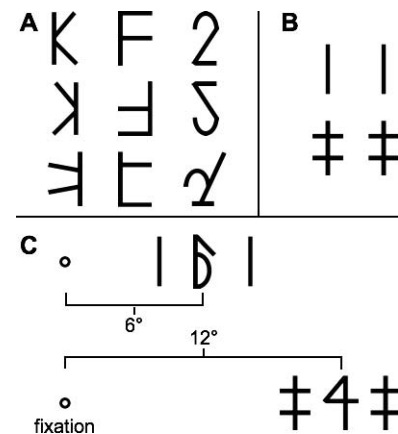


Figure 1. Stimulus examples of Experiment 1. (A) Targets consisted of three basic line elements (either three lines, or an arc and two lines), making up numbers, letters, meaningless symbols, and rotated numbers and letters (nine target examples are shown). (B) There were two types of flankers: Simple flankers (low crowding condition) consisted of a single long line; complex flankers (high crowding condition) of one long and two short lines. (C) Stimuli were presented in the right visual field at 6° and 12° eccentricity (and with free viewing in the center). Stimuli are not drawn to scale.

were about 0.7° wide (with the exception of the simple flankers, and some variance depending on the item). The center-to-center spacing between target and flankers was 0.8° . When observers did not keep fixation, stimuli were masked by a pattern mask consisting of an arrangement of the three basic elements (long bars, short bars, arcs), presented at the target and flanker locations. All elements were black with a luminance of 0.1 cd/m^2 , presented on a gray background (50.5 cd/m^2).

Design and procedure

Observers fixated on a fixation dot in the center of the screen. Stimuli were only presented when observers fixated on the dot. When observers did not maintain fixation, the mask was presented at the stimulus location. A feedback tone was given when observers erroneously made an eye movement towards the stimulus. Participants were asked to draw each stimulus into a rectangular area indicated by faint corners in the drawing book. We instructed participants to draw the stimulus as it appeared to them, making the drawing with free viewing resemble the peripherally viewed stimulus as much as possible. The entire stimulus, i.e., target and flankers, was drawn.

On each trial, observers initiated the stimulus presentation by fixating the fixation dot. There were no restrictions regarding the number and duration of fixations on a given trial. Drawings were made with free viewing. Observers shifted their gaze between the

fixation dot on the screen and the drawing board, without leaving the head rest. Participants indicated verbally when they finished a drawing, and the experimenter initiated the next trial. The eye tracker was recalibrated when necessary.

Stimuli were presented at 0°, 6°, and 12° eccentricity. Observers completed two blocks with 86 trials per block, consisting of 72 trials in which targets were flanked by simple (36 trials) or complex (36 trials) flankers. In the 14 remaining trials, randomly selected targets were presented without flankers. In half of the trials, the target was presented at 6° eccentricity, in the other half at 12° eccentricity. All stimulus conditions were intermixed within a block, and stimuli were presented in random order. In the second block, the stimulus order was reversed. In four additional trials, stimuli were presented at the foveal location (with free viewing) to confirm that participants had basic drawing skills.

Scoring

To analyze the drawings, we quantified the differences between a presented stimulus and the corresponding drawing using the scoring system described below. The goal was to capture perceptual deviations or “errors” when stimuli were crowded. For that aim, we suggest a number of error classes and error types that can be used to describe the differences between the presented stimuli and the drawings. There are infinite ways of analyzing the transformations between a stimulus and the drawings. The error types proposed here are motivated by spatial transformations on dimensions reported in earlier crowding studies, such as feature location, orientation, and substitution, as well as phenomenological reports describing characteristics of appearance in crowding (e.g., Korte, 1923). Two scorers were trained in scoring the errors. They were naïve as to the purpose of the experiment, and evaluated each drawing in regard to four element error classes: number, size, position, and shape (see Table 1). Elements were defined as continuous lines of a minimum size that allowed clear identification of their orientation (continuous lines consisting of two or more pen strokes were defined as a single element). The number of elements of each error type was rated. In the Number class, scorers indicated the number of Omissions and Additions of elements. In the Size class, each depicted element was classified as shorter (Truncation), longer (Extension), or equal as in the stimulus. Size scorings were made relative to the depicted extent of the drawing, indicating deviations of at least 30% compared to correctly depicted elements, allowing for unintentional drawing inaccuracies. In the (linear and angular) Position class, the position of an element relative to the other depicted elements was classified as

accurate or inaccurate. A Translation error was rated for minimum deviations of at least 30% of the length of the elements. Rotation deviations of more than 15° were rated as errors. In the Shape class, elements were rated as correct or distorted, for example, when curved in the drawing but straight in the stimulus (Distortion). Scorings were made using the smallest number of required changes to match a drawing to the original stimulus. For example, when an element was depicted in the wrong position, a Translation was rated, not an Omission of an element at that position and an Addition in the new (wrong) position. Therefore, in the Number class, elements were never added and missing within a single item. However, as elements could be perceived at a wrong location, for example, a flanker element at the target location, we analyzed the relation of Omission and Addition errors in the target and the flankers to quantify potential migrations of elements between items.

Note that the number of presented elements determines the maximum number of errors (for all error types except for Addition errors). As there were three target elements, the dynamic range of each error type was between zero and three (except for Addition). No magnitudes of deviations were scored. Besides target element errors, target-flanker substitutions were scored when a flanker was depicted at the target position and/or the target at a flanker position.

Results and discussion

The main results are reported as average error rates, calculated by dividing the number of errors of each error type by the number of scored stimuli. For example, the resulting quotient is 1, if each scorer notes one error per stimulus. If on average more than one error occurs per stimulus, the quotient is larger than 1; if less than one error occurs, it is smaller than 1. To evaluate interscorer agreement, we calculated Cohen’s kappa coefficients for each error type (Cohen, 1960). Interscorer agreement was high. The average kappa for the seven element error types was 0.89 (Translation: 0.82, Truncation: 0.83, Rotation: 0.87, Omission: 0.90, Addition: 0.93, Extension: 0.93, Distortion: 0.94).

To analyze crowding errors independent of peripheral visual resolution and self-crowding (“internal crowding,” Martelli, Majaj, & Pelli, 2005; “within-character” crowding, Zhang, Zhang, Liu, & Yu, 2009), we subtracted the baseline error rates from the crowded error rates for each error type. Figure 2 shows the crowded error rates (bars) and the baseline error rates (dashed lines). First, we conducted a two-way MANOVA with the factors Eccentricity (6° and 12°) and Flanker Complexity (simple and complex), and the error rates of the seven element error types as

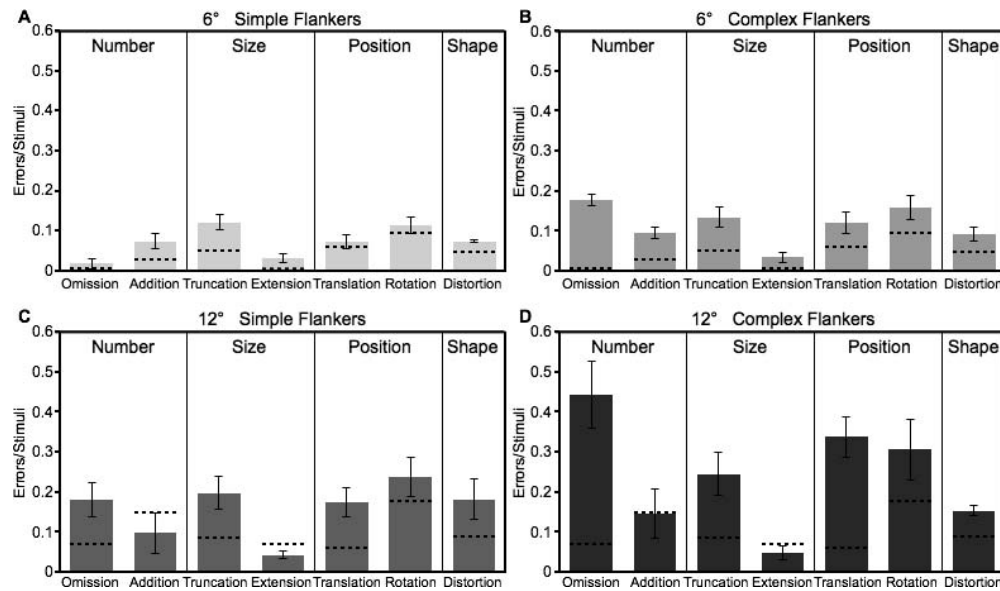


Figure 2. Target element error rates in Experiment 1. The average number of errors is shown for each error type for targets presented at 6° (A, B) and 12° (C, D) eccentricity. Targets were flanked by simple (A, C) or complex (B, D) flankers. Normalized error rates were obtained by subtraction of the baseline error rates—shown by the dashed horizontal lines—from the error rates in the conditions with flankers. Error bars indicate standard errors of the mean.

dependent variables. Error rates were higher when stimuli were presented at 12° eccentricity compared to 6°, $F(7, 10) = 10.72$, $p < 0.005$, $\eta_p^2 = 0.88$. Subsequent ANOVAs revealed that Omission, $F(1, 16) = 5.09$, $p < 0.05$, and Truncation, $F(1, 16) = 31.09$, $p < 0.001$, error rates were higher at 12° than at 6°. There was also a significant difference for Extension errors, $F(1, 16) = 4.94$, $p < 0.05$, however, with a higher error rate at 6° than at 12°. There were no differences between the remaining error types at 6° compared to 12° eccentricity. The MANOVA also revealed higher error rates with complex than with simple flankers, $F(1, 16) = 11.53$, $p < 0.001$, $\eta_p^2 = 0.89$. Subsequent ANOVAs showed that the Omission error rate, $F(1, 16) = 10.93$, $p < 0.005$, and the Truncation error rate, $F(1, 16) = 13.23$, $p < 0.005$, were higher in the complex than in the simple flankers condition. The other error types did not differ in the two flanker conditions. There was no interaction between Eccentricity and Complexity, $F(7, 10) = 1.64$, $p = 0.23$.

Our central question was to what extent the different error types occur in crowding. To compare the seven error types, we first conducted a repeated-measures ANOVA with the seven error types as additional factor. Sphericity was tested with Mauchly's sphericity test; Greenhouse-Geisser or Huynh-Feldt corrections were applied when required. The ANOVA revealed differences between the error types, $F(2.4, 9.7) = 4.32$, $p < 0.05$, $\eta_p^2 = 0.52$ (Greenhouse-Geisser corrected); an interaction between error type and flanker complexity, $F(1.8, 7.2) = 7.83$, $p < 0.05$, $\eta_p^2 = 0.66$ (Greenhouse-Geisser corrected); and a trend for an interaction

between error type and eccentricity, $F(6, 24) = 2.45$, $p = 0.54$, $\eta_p^2 = 0.38$ (Huynh-Feldt corrected). Replicating the MANOVA results, the ANOVA also showed the main effects of eccentricity, $F(1, 4) = 18.30$, $p < 0.05$, $\eta_p^2 = 0.82$, and flanker complexity, $F(1, 4) = 44.07$, $p < 0.005$, $\eta_p^2 = 0.92$, and no interaction between eccentricity and complexity, $F(1, 4) = 4.30$, $p = 0.11$, $\eta_p^2 = 0.518$. There was no three-way interaction, $F(1.7, 6.8) = 1.22$, $p = 0.34$, $\eta_p^2 = 0.23$ (Greenhouse-Geisser corrected).

Following these results, we compared each two error types within each of the four eccentricity \times flanker complexity conditions using Tukey tests for pairwise comparisons. In the complex flanker condition at 12° (the condition with the expected strongest crowding), the Omission error rate was higher than the Addition ($p < 0.01$), Extension ($p < 0.01$), and Distortion ($p < 0.05$) error rates. The Translation error rate was higher than the Addition ($p < 0.05$), and Extension ($p < 0.05$) error rates. In the complex flanker condition at 6°, the Omission error rate was higher than the error rates of all other error types (all $p < 0.01$, except for Addition and Truncation: $p < 0.05$). Comparisons between the other error types in the complex flankers conditions, and all comparisons in the simple flankers conditions did not yield any differences; however, several comparisons showed trends for differences. To further characterize the error rate distributions, we calculated effect sizes (Cohen's d ; Cohen, 1988) for comparisons of each two error types in all four eccentricity \times flanker complexity conditions (using the pooled standard deviation of

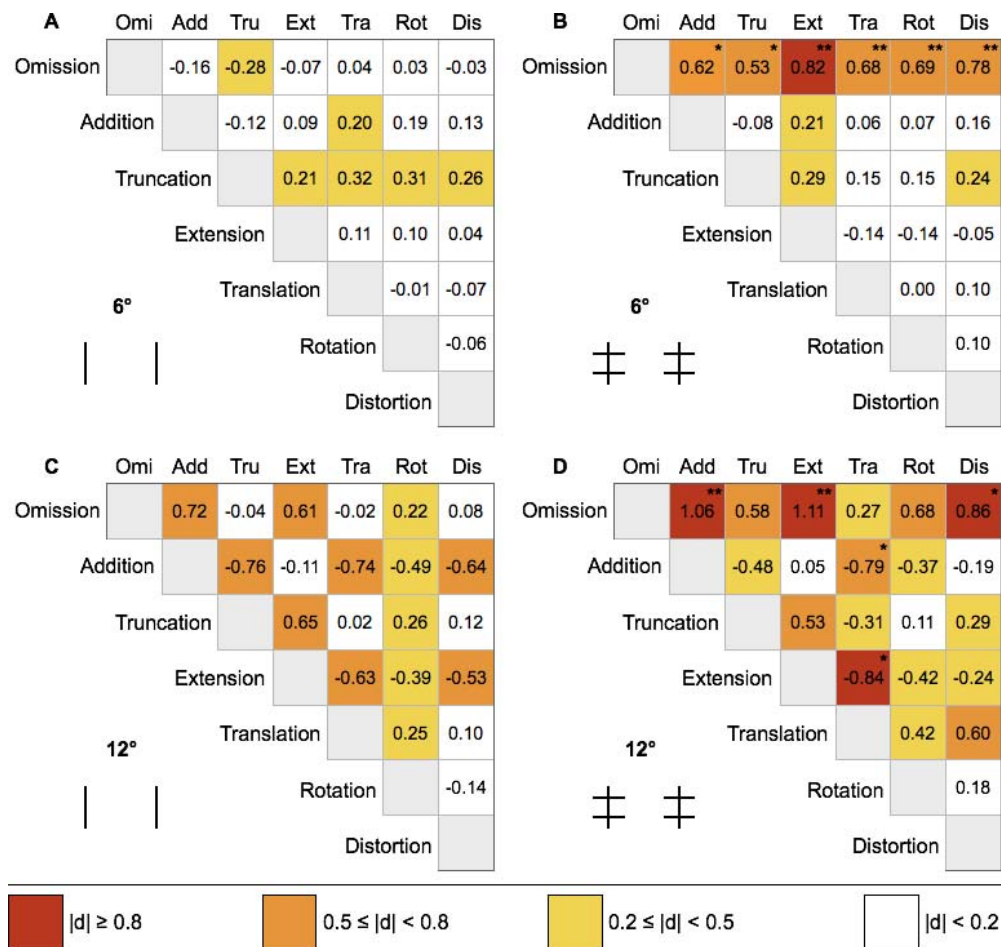


Figure 3. Cohen's d for comparisons of each two error types at 6° eccentricity for simple (A) and complex (B) flankers, and at 12° eccentricity for simple (C) and complex (D) flankers in Experiment 1. Colors indicate large (red), medium to large (orange), small to medium (yellow), and small (white) effect sizes, following Cohen's (1988) convention for effect size categories. Cohen's d is shown for subtractions of error rates in the rows from those in the columns, i.e., positive values indicate higher error rates in rows. Asterisks indicate significance of pairwise comparisons.

each of the four conditions for standardization). The resulting effect sizes are shown in Figure 3.

As shown in Figures 2 and 3, the Omission error rate was particularly high compared to the other error rates. Because omitted target elements could be missing completely or be drawn at a flanker location, we next analyzed to what degree target element Omissions were balanced by flanker element Additions. For each target stimulus in which at least one element was missing, we subtracted the additionally drawn elements in each of the flankers from the number of missing target elements. If target elements "migrated" to a flanker position, we would expect large proportions of balanced missing target elements and added flanker elements. We found that overall in 11.1% of the cases, missing target elements were balanced by additional flanker elements (3.5% in flanker 1, 6.0% in flanker 2, and 1.6% in both flankers). In 88.9% this was not the case. The same analysis for flankers revealed a similar result: Only 5.8% of the Omission errors in the flankers

(3.4% in flanker 1, and 2.4% in flanker 2) were balanced by added target elements. Addition errors, however, were balanced by Omissions in 22.5% when they occurred in the target (10.5% in flanker 1, 7.5% in flanker 2, and 4.5% in both flankers) and 25.2% (9.8% in flanker 1, and 15.4% in flanker 2) when they occurred in the flankers.

Besides element errors, we investigated the rate of target-flanker substitutions, i.e., the frequency of entire targets drawn at a flanker location and/or vice versa. Almost no complete target-flanker substitutions were observed (error rate: 0.01).

Taken together, our results show clear crowding errors. Error rates were higher at 12° compared to 6° eccentricity, and higher with complex than simple flankers. The main effect of eccentricity was driven by the Omission and Truncation error rates which were higher at 12° eccentricity, indicating that these error types were particularly characteristic of crowding. Interestingly, Extension errors were higher at 6° than at

12° eccentricity, possibly due to the relatively high Extension error rate in the unflanked condition at 12° compared to 6° eccentricity. Note, however, that the Extension error rates were low in all conditions, indicating that Extension errors did not depend on crowding. The two flanker conditions (simple and complex) were used to systematically vary crowding strength. Compared to the complex flankers, the simple flankers were expected to cause only weak crowding as they strongly differed from the targets on several dimensions. The higher Omission and Truncation error rates in the complex compared to the simple flankers condition indicate again that these error types were genuine crowding errors.

Comparisons of the error types revealed pronounced error rate differences, particularly in the condition with the expected strongest crowding, i.e., complex flankers at 12°, as well as the complex flankers condition at 6° eccentricity. Before discussing the specific error rate differences, it is important to note that comparisons across error classes, for example between Number and Size errors, are subject to the general limitations of comparing different stimulus dimensions. While our comparisons are based on a common unit, i.e., error rates, the definition of errors is dimension-specific. For example, Number errors are based on changes of the quantity of elements, and Size errors are based on changes of the length of elements. Hence, error rate comparisons between classes are dependent on what determines an error in each class. Similar constraints hold for the within-class comparison of Position errors, where Translation and Rotation errors do not share a common dimension. Comparisons within the Number and Size error classes, i.e., Omission compared to Addition errors, and Truncation compared to Extension errors, however, are not subject to the same dimensional constraints because each two error types within a class are determined on the same dimension.

We expected to find differences between error types in conditions with strong crowding as crowding-specific errors should increase with crowding strength whereas errors not specific to crowding should not increase. Within-class comparisons revealed that Omission error rates were higher than Addition error rates in the two complex flankers conditions (6° and 12° eccentricity). This finding indicates that Number errors are not equally distributed in crowding but rather, that observers tend to perceive less elements than were presented. While the Truncation error rate was higher in the high compared to the low crowding conditions (12° vs. 6° and complex vs. simple flankers), the within-class comparison with Extension errors did not reach significance. However, there was a trend for higher Truncation than Extension error rates in all conditions (see also Cohen's d in Figure 3). We suggest that these within-error-class comparisons indicate that crowding

caused “*diminishment*,” a decrease in the perceived number and possibly size of elements.

Comparisons across error classes showed that in the complex flankers conditions, Omission error rates were higher than the error rates of all other error types at 6° eccentricity, and higher than the Extension and Distortion error rates at 12° eccentricity. The comparably high Omission error rates in high (but not in low) crowding conditions indicate that Omission errors were due to crowding. At 12° eccentricity, the relatively high Translation error rate possibly reflected the well-known location uncertainty in crowding. Rotation errors, on the other hand, occurred less frequently, and—as Extension, Addition, and Distortion errors—did not vary strongly with crowding strength. The absence of differences between error types in the simple flankers conditions is in line with the expected low crowding strength in these conditions. However, particularly at 12°, there were some trends for differences between the error types, such as a relatively high Omission compared to Addition error rate (see also Figure 3).

The prominence of target element omissions under high crowding could be due to element migrations from the target to the flankers. However, in the large majority of trials, this was not the case, showing that element migration did not underlie the effect of omitted target elements. Instead, target elements were largely omitted without being redrawn at one of the flanker locations. The same holds for Omission errors in the flankers. Addition errors, however, were balanced by Omission errors in about a fifth to a quarter of the cases, indicating a higher probability of feature migration. Note that not all possible feature migrations can be captured by the drawing method. For example, perceptually swapping a vertical target element with a horizontal flanker element would be scored as two Rotation errors—one in the target and one in the flanker—and not as feature migration. Importantly, the substitution of entire targets/flankers did not play a role in our results.

Some error rates were also relatively high in the unflanked condition which was possibly due to self-crowding which is expected when using complex targets with multiple features (Martelli et al., 2005; Zhang, Zhang, Xue, Liu, & Yu, 2009; Coates, Wagemans, & Sayim, 2017). Drawings of stimuli presented at 0° resulted in target error rates below 0.05 for each error type (except for a *Distortion* rate of 0.11). Hence, limited drawing abilities and the drawing method in general did not determine our results. However, there was a relatively low correlation of error rates between the first and the second presentation of the same stimulus which could be related to aspects of the drawing method (Appendix A).

All in all, the results indicate that Omission and Truncation errors were most characteristic of crowding

in this experiment. Importantly, Omission errors were also related to identification performance. In an additional experiment (see Appendix B), we asked a different set of observers to indicate the identity of the target drawings in a free-naming task, and found that Omission errors were relatively strongly correlated with performance ($\rho = -0.40$): The more Omission errors were made in the initial drawings, the harder it was to identify the target. Hence, Omission errors were related to a standard measure of crowding as well.

Experiment 2

Because only two different flanker types were presented in Experiment 1, it is likely that—due to familiarization—observers were not maximally naïve about the presented stimuli (despite the high variance of target identities). To minimize the effect of flanker familiarity, we varied flanker identities in Experiment 2. Furthermore, stimulus knowledge obtained in the presentations at 6° potentially influenced the perception of stimuli at 12° eccentricity in Experiment 1. To reduce possible influences of prior knowledge, Experiment 2 always started with stimulus presentation at 12° followed by 6° . Unlike in Experiment 1, we used only meaningless line configurations to keep participants as uninformed as possible about all aspects of the stimuli. As potential eccentricity differences cannot be analyzed independent of presentation order, and target-flanker similarity was approximately kept constant, there was no systematic variation of crowding strength.

Methods

Experiment 2 was the same as Experiment 1 except for the following. Stimuli consisted of pseudo-random line configurations, created by using similar elements as in Experiment 1 (see Figure 4). Neither the target nor any of the flankers was the same within or across stimuli, hence, each target-flanker configuration was unique. As in Experiment 1, observers were asked to draw the entire stimulus (target and flankers). We use the terms “Flanker 1,” “Target,” and “Flanker 2” to designate the three items from left to the right (close to fixation to far from fixation). Observers were not informed about the structure of the stimuli.

Experiments started with the condition at 12° eccentricity, followed by 6° , and last the foveal condition, to keep observers as naïve about the stimuli as possible. Observers completed five blocks of trials—two blocks at 12° , two blocks at 6° and one block at 0° eccentricity. In the first block, 20 stimuli consisting of target and flankers, and four targets without flankers

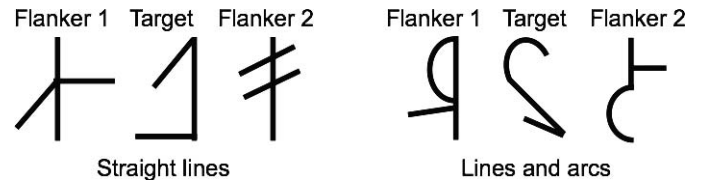


Figure 4. Stimulus examples of Experiment 2. Each item (Flanker 1, Target, and Flanker 2) consisted either of straight lines, or lines and an arc. Stimuli were presented in the right visual field. The entire stimulus (target and flankers) was drawn.

were presented in a random order. In the second block, the stimulus order was reversed. At 0° , six stimuli consisting of a target and two flankers were presented. Four new observers participated in Experiment 2. One additional observer did not finish the experiment and was excluded.

Results and discussion

As in Experiment 1, two scorers evaluated each drawing in regard to the seven element error types and target-flanker substitutions. The element errors were rated individually for each position (Flanker 1, Target, Flanker 2). Interscorer agreement was substantial. The average kappa for the seven element error types was 0.77 (Omission: 0.70, Rotation: 0.70, Translation: 0.70, Addition: 0.74, Truncation: 0.79, Extension: 0.85, Distortion: 0.89).

The element error distributions of the seven error types for targets and flankers are shown in Figure 5. After subtracting the baseline error rates from the flanked target error rates for each error type, we conducted a repeated-measures ANOVA with the two factors Error Type and Eccentricity. Sphericity was tested with Mauchly’s sphericity test; Greenhouse-Geisser or Huynh-Feldt corrections were applied when required. Flanker errors were not included in the analysis because baselines were only measured at the two target locations. The results showed that there were differences between the target error rates, $F(6, 18) = 4.59$, $p < 0.01$, $\eta_p^2 = 0.61$. There was no effect of Eccentricity, $F(1, 3) = 0.21$, $p = 0.68$, $\eta_p^2 = 0.06$ (Greenhouse-Geisser corrected). However, we found a significant interaction between Error Type and Eccentricity, $F(6, 18) = 4.35$, $p < 0.01$, $\eta_p^2 = 0.59$. Next, we used Tukey tests for pairwise comparisons to compare each two error types within each of the two eccentricity conditions, and calculated effect sizes (Cohen’s d , Figure 6) for comparisons between each two error types. Comparisons within error classes showed that at 12° eccentricity, the Omission error rate was higher than the Addition error rate ($p < 0.01$). The other comparisons within error classes did not reach significance; however, there were trends for a higher

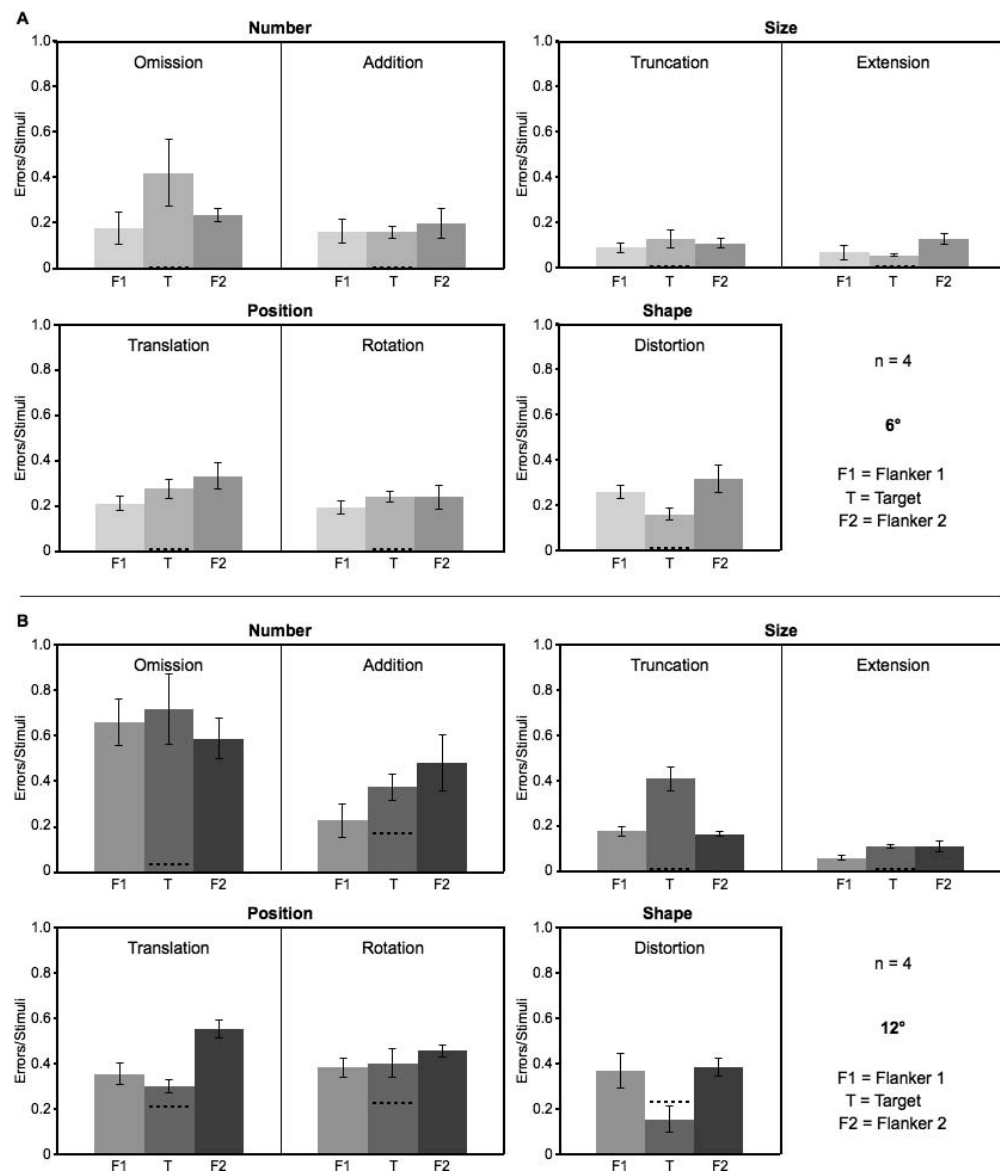


Figure 5. Element errors in Experiment 2. The bars show the average number of errors of each error type in the four error categories Number (Omission and Addition), Size (Truncation and Extension), Position (Translation and Rotation), and Shape (Distortion). Each bar triplet shows from left to right errors for flanker 1, the target, and flanker 2. See text for details. The dashed lines show unflanked error rates at the target location. (A) Error rates at 6° eccentricity. (B) Error rates at 12° eccentricity. Error bars indicate standard errors of the mean. Note the difference in scale compared to Experiment 1.

Omission than Addition error rate at 6° , and a higher Truncation than Extension error rate at 12° eccentricity. Comparisons across error classes revealed that at 12° eccentricity, the Omission error rate was higher than the error rates of all other error types (Extension, Translation, Distortion: $p < 0.01$; Addition and Rotation: $p < 0.05$), except for Truncation errors. Whereas the remaining comparisons at 12° did not show any differences, there were strong trends for a higher Truncation error rate compared to the Distortion and Translation error rates. At 6° eccentricity, the Omission error rate was higher than the Truncation ($p < 0.05$) and the Extension ($p < 0.01$) error rates. There

was also a trend for a higher Omission error rate compared to the Addition and Distortion error rates (see also Figure 6).

As in Experiment 1, the target's Omission error rate was high. To test how far feature migration to the flankers might underlie this result, we asked again if missing target elements were balanced by added flanker elements. In 31.7% of the cases, target Omissions were balanced by flanker Additions (10.7% in Flanker 1, 18.3% in Flanker 2, and 2.7% by both flankers). Omissions of flanker elements, on the other hand, were balanced by added target elements in only 14.9% (5.9% in each flanker, and 3.0% in both flankers). Addition

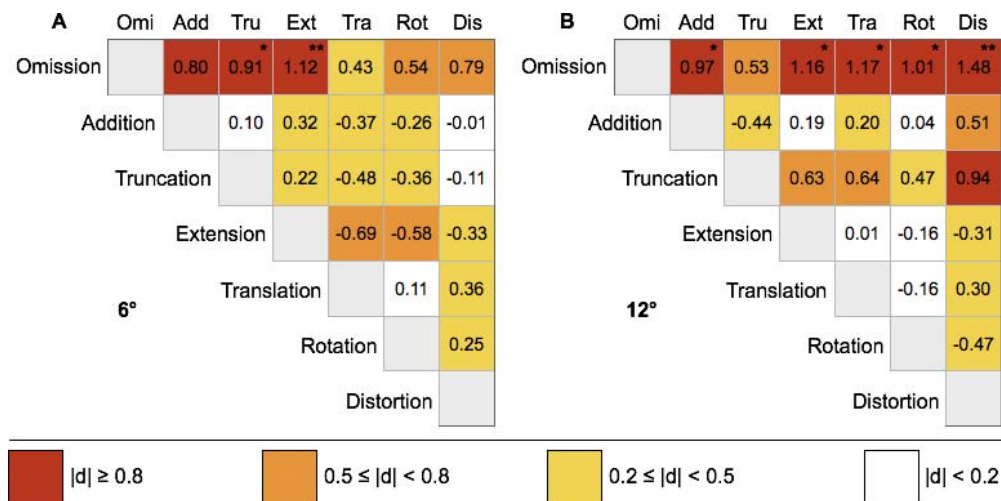


Figure 6. Cohen's d for comparisons of each two error types at 6° (A) and 12° (B) eccentricity in Experiment 2. Colors indicate large (red), medium to large (orange), small to medium (yellow), and small (white) values, following the notation of Cohen (1988). Cohen's d is shown for subtractions of error rates in the rows from those in the columns, i.e., positive values indicate higher error rates in rows. Asterisks indicate significance of pairwise comparisons.

errors were balanced in 57.4% (17.8% Flanker 1, 14.9% Flanker 2, 24.8% both flankers) when occurring in the target and 41.6% (17.9% Flanker 1, 22.6% Flanker 2, 1.1% both flankers) when occurring in the flankers.

To explore if the element error rates differed between Flanker 1, the Target, and Flanker 2, we compared the error rates of the three items using the original, not baseline-corrected values (because no baselines were measured at the flanker locations). The data were analyzed with a two-way MANOVA with the factors Item Position (flanker 1, target, and flanker 2) and Eccentricity (6° and 12°), and the seven Error Types as dependent variables. The MANOVA showed that the error rates differed between the three items (main effect of Item Position; $F(14, 26) = 3.56$, $p < 0.005$, $\eta_p^2 = 0.66$). Follow-up ANOVAs for each error type showed Item Position differences for Truncation: $F(2, 18) = 11.31$, $p < 0.005$; Extension: $F(2, 18) = 4.02$, $p < 0.05$, Translation: $F(2, 18) = 9.40$, $p < 0.005$; and Shape errors: $F(2, 18) = 8.11$, $p < 0.005$. Comparisons of the three items for each of these error types using Tukey tests revealed that the target had a higher Truncation error rate (both $p < 0.005$) and a lower Distortion error rate than each of the two flankers (Flanker 1: $p < 0.05$, Flanker 2: $p < 0.005$). Flanker 2 had a higher Extension error rate than Flanker 1 ($p < 0.05$), and a higher Translation error rate than Flanker 1 and the target (both $p < 0.005$). The MANOVA also showed a significant effect of Eccentricity, with higher error rates at 12° compared to 6°, $F(7, 12) = 7.16$, $p < 0.005$, $\eta_p^2 = 0.81$. Subsequent separate ANOVAs for each error type revealed that errors at 12° were higher compared to 6° eccentricity: Omission, $F(1, 18) = 18.20$, $p < 0.001$; Addition, $F(1, 18) = 10.37$, $p < 0.01$; Truncation, $F(1, 18) = 31.30$, $p < 0.001$; Translation, $F(1, 18) = 13.20$, p

< 0.005 ; and Rotation, $F(1, 18) = 31.30$, $p < 0.001$. There was no interaction between Item Position and Eccentricity, $F(14, 26) = 1.61$, $p = 0.144$, $\eta_p^2 = 0.46$.

Besides element errors, we evaluated target-flanker substitutions. Almost no complete target-flanker substitutions were observed. The average substitution error rate was 0.04. Hence, as in Experiment 1, substitution of entire items did not play a role in Experiment 2.

The results of Experiment 2 showed clear differences between error types and a similar pattern of results as in Experiment 1. In particular, the higher Omission compared to Addition error rate shows that observers tended to perceive less elements, indicating that target diminishment played a key role. The trend for higher Truncation than Extension error rates at 12° eccentricity, i.e., when crowding was strong, supports the notion of target diminishment. The comparisons across error classes at 12° eccentricity highlighted the prominence of Omission errors which were more frequent than all other error types (see also Figure 6).

Interestingly, as about 1/3 of target element Omissions were balanced by flanker element Additions, a comparatively large part of Omission errors could be due to feature migration from the target to the flankers. Migration from the flankers to the target, however, was less likely. As expected, the low number of Addition errors was frequently balanced by Omissions, especially in the target. Except for the balancing of Addition errors in the target, the majority of Omission and Addition errors, however, were not balanced, indicating that even though feature migration could have played a role, it cannot underlie most Omission and Addition errors. In general, the higher level of balancing of Addition errors compared to Omission errors highlights the prominence of Omissions because

a larger number of Addition errors could be due to feature migration. The rather high balancing rates compared to Experiment 1 were possibly observed because flankers were not fixed to one of two types but varied strongly.

Not only the target but also the flankers were subject to crowding. Comparing the (not normalised) error rates at the three locations revealed some differences between the items. We expected that target error rates would be higher than flanker error rates as the target was flanked on both sides whereas the two flankers were only flanked on one side. However, this was only the case for Truncation errors (and the opposite for Distortions), and overall the error rates were similar for the three items. Importantly, not only the number of adjacent items (2 for the target, 1 for each flanker), but also the location of the items (inward, outward, or both), and the eccentricity differed between the items. Hence, the observed differences are due to a combination of (at least) three variables, and cannot be conclusively disentangled. The higher Extension and Translation error rates of Flanker 2, however, may well be due to its location farther in the periphery.

Comparing the two eccentricity conditions revealed higher error rates at 12° than 6° eccentricity, but no difference when the normalized data was compared. Because of the relatively small dynamic error range, the different baselines in the two eccentricity conditions could have obscured the effect with normalized data. As there was a fixed order of eccentricities (first 12° then 6°), however, the two Eccentricity conditions are not independent, and were mainly compared for explorative reasons.

In contrast to Experiment 1 where only two different flanker types were presented, the flankers were different in each stimulus in Experiment 2. Similar element error distributions in Experiment 1 and 2, including the predominance of Omission errors, show that flanker familiarity did not strongly influence error rates. While flanker familiarity could have also caused the lack of substitution errors in Experiment 1, the equally low substitution error rate in Experiment 2 suggests that flanker familiarity does not account for the lack of substitution errors. Rather, we suggest that multiple views of the same stimulus during a single trial might underlie the low substitution rates (what needs to be validated by future studies).

General discussion

Most crowding paradigms measure performance. However, crowding does not only deteriorate performance but also alters appearance. While the phenomenology of crowded, peripheral vision has been

described in detail already early in the history of crowding research (Korte, 1923; for a summary of Korte's work, see Strasburger, 2014), systematic investigations of appearance to understand the underlying mechanisms of crowding have only recently started (Greenwood et al., 2010; Coates, Wagemans, Sayim, 2017; Koenderink et al., 2017). Here, our goal was to characterize crowded, peripheral appearance by assessing what stimulus features survive crowding and to what degree features are changed, added, or lost in crowding. For that aim, we introduced an appearance-based drawing method to capture what is perceived in the crowded periphery (see also, Sayim, Myin, & Van Uytven, 2015; Coates, Wagemans, & Sayim, 2017), and proposed a range of error types that describe the transformations between presented stimuli and representations of their peripheral appearance. To quantify the transformations, each drawing was rated in regard to the number, size, position, and shape of its elements compared to the presented stimulus, yielding characteristic error type distributions. Additionally, we asked if entire items were mislocalized, that is, if targets were substituted with flankers and vice versa.

We found clear differences between error types. Before discussing the error distributions in more detail, we first outline the usefulness of the developed drawing method to capture appearance changes and errors in crowding. In most crowding studies, responses are restricted by providing observers with only few response alternatives (alternative forced choice methods). Therefore, detailed characteristics of the appearance of crowded targets are usually lost. Drawing has the advantage in that there are much fewer response restrictions compared to other response formats, and hence, appearance changes and a broad range of error types can be captured that may not be noticed in standard crowding paradigms. For example, when measuring performance with predefined response categories, errors such as omissions of entire elements are not detected because observers respond with what fits the available response categories best. Indicating whether a target letter T is presented upright or upside down, for example, can be performed without ever perceiving the vertical line of the T by judging the relative position of the horizontal line. Given the response categories “upright” or “upside down,” aspects of the actual appearance of the stimulus would not be available in the data. The drawing method prevents such response format driven losses of appearance information, and enables a more fine-grained typology of crowding errors compared to other classifications (e.g., Huckauf & Heller, 2002; Freeman, Chakravarthi, & Pelli, 2012; Hanus & Vul, 2013).

Unconstrained verbal descriptions of a stimulus can potentially yield similarly detailed accounts of appearances as drawings (see also Metzger, 1936). However,

verbal descriptions of a percept do not match the precision of drawings as soon as stimuli reach a certain level of complexity. For example, capturing the percept of a random arrangement of multiple lines by drawing is more precise and efficient than verbally describing the relative locations, orientations, and junctions, etc. Another way to increase information about appearance in crowding is to use more precise response alternatives. For example, presenting a large number of response alternatives (e.g., letters: Wang, He, & Legge, 2014; digits: Strasburger & Malania, 2013; Chinese characters: Zhang, Zhang, Liu, & Yu, 2012), or relating observers' responses to the similarity among the response items (Bernard & Chung, 2011) may give a better account of the appearance of complex targets than paradigms with more restricted response formats.

Our drawing results show that several different error types can be characterized in crowding, and that some of these error types were more common than others. In the Number class, Omission errors were clearly more common than Addition errors, suggesting that elements were suppressed (Krumhansl & Thomas, 1977; Chastain, 1981), or lost at early (Balas et al., 2009; Rosenholtz, Huang, & Ehinger, 2012) or later (He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001; Petrov & Popple, 2007) stages of visual processing. We can exclude that the high Omission error rates were an artifact of the drawing method or the rating procedure, as Omission errors strongly depended on crowding strength. Error rates were higher with complex than with simple flankers and at 12° compared to 6° (Experiment 1). This was not the case for the other error types, except for Truncation errors. Whereas comparisons across error classes are more constrained than within error classes, and depend more strongly on the definition of what counts as an error, it is nevertheless worth noting that Omission error rates were high compared to the other error types in most conditions, particularly when crowding was strong. Importantly, Omission errors were also correlated with identification performance and therefore related to a standard measure of crowding (Appendix B).

The only additional error type that showed clear dependence on crowding strength was Truncation. Stronger crowding yielded higher Truncation error rates, indicating that Truncation errors—as Omission errors—were genuine crowding errors. The Extension error rates were low in all conditions, and therefore it seems that they were not characteristic of crowding. Surprisingly, although Translation error rates were relatively high, they did not increase with crowding strength even though position uncertainty is one of the prominent characteristics of crowding, occurring on the level of features (Wolford, 1975) as well as entire items (Estes et al., 1976). In general, due to location

uncertainty in peripheral vision (Levi, Klein, & Yap, 1987), Translation errors are also expected under uncrowded conditions which was partly reflected in the unflanked Translation error rates. While Distortion errors reflect the often reported phenomenology of crowding, they did not occur as frequently as expected. We suggest that stimuli consisting of only straight or curved lines, and a ballpoint pen to draw, limited the rate of Distortion errors compared to what would be expected when presenting more complex stimuli and providing more flexible drawing tools (Sayim, Myin, & Van Uytven, 2015).

The reported Translation errors occurred within a single item. However, feature mislocalization and translation can also occur between items (feature migration; Wolford, 1975; Wolford & Shum, 1980; Nandy & Tjan, 2007). The drawing paradigm cannot distinguish with certainty between such feature migration between items and Omission/Addition errors; however, it is possible to quantify the rate of the *potential* migrations of elements between the target and the flankers by comparing Omission and Addition errors in the target and the flankers. Our analysis revealed that in most trials, Omission errors in the target and the flankers were not balanced by added elements. Hence, feature migration may only partly explain Omission errors—the majority of Omission errors must have been caused by a different process. In comparison, additional elements in the targets and the flankers were more frequently balanced by omitted elements (especially in Experiment 2), making it more likely that Addition errors were in some cases caused by element migration.

Mislocalizations occur in crowding not only on a feature level but also on the level of entire items (substitution; e.g., Strasburger et al., 1991; Chung & Legge, 2009). In our experiments, however, the substitution of the target by a flanker or vice versa was extremely rare. To test whether the low substitution rate in Experiment 1 was driven by the presentation of only two different flanker types, we varied the flankers in each stimulus in Experiment 2. Again almost no target-flanker substitutions were observed. A possible explanation for the small rate of substitutions is that stimuli were presented with unlimited viewing time. For example, it was shown that substitution and error rates were strongly reduced with increasing presentation times (Styles & Allport, 1986). Other studies, however, showed that unlimited viewing time only marginally changed crowding (Wallace, Chiu, Nandy, & Tjan, 2013), and high substitution rates also occurred with relatively long presentation times (2.4 s, Estes, Allmeyer, Reder, 1976; 1.6 s, Zhang et al., 2012). An alternative explanation is that multiple presentations, i.e., multiple (peripheral) views of a single stimulus during one trial, caused the observed low substitution

rates. As substitution does not occur on every trial and not necessarily with the same features or items (Ester, Klee, & Awh, 2014), mislocalizations of items that occurred during one view of the stimulus could possibly be counteracted by further views of the same stimulus during the same trial.

To conclude, we have introduced an appearance-based drawing method and a categorization system that can be used to characterize appearance changes and quantify phenomenological aspects of crowding. In contrast to earlier classifications of crowding errors, we propose a more fine-grained typology. Our results suggest that in addition to the loss of positional information, target diminishment in the form of Omissions and Truncations of elements are a key feature of crowding. As we did not quantify statistical differences between the drawings and the stimuli, it remains to be shown if the information loss we report here differs from the information loss that occurs when extracting summary statistics as proposed by Rosenholtz and colleagues (Balas et al., 2009; Rosenholtz et al., 2012). Entire target-flanker substitutions did not play a role in our results; however, the balancing of element Omission and Addition errors in the target and the flankers indicate that feature migration did contribute to the error distributions. Importantly, most of the reported error types are not distinguished in standard crowding paradigms in which performance is measured by using only few response alternatives. We propose that the error types suggested here are a useful basis for future studies to further quantify error characteristics in crowding. The specific contributions of response format, stimulus type, and presentation condition (e.g., multiple consecutive views of the same stimulus) will need further exploration in future studies. We suggest that drawing is a valuable method to investigate appearance in crowding and other perceptual phenomena that are not entirely captured by verbal descriptions or simple discrimination tasks.

Keywords: crowding, appearance, drawing, phenomenology, appearance-based methods

Acknowledgments

We thank the Ecole Supérieure des Arts (ESA) Saint-Luc in Brussels and all students who participated in the research. We thank Daniel Coates for helpful discussions of earlier versions of the manuscript. This work was supported by an FWO Pegasus Marie Curie fellowship awarded to BS, the Swiss National Science Foundation (PP00P1_163723/1 awarded to BS), and the Methusalem program by the Flemish Government (METH/08/02 and METH/14/02) awarded to JW.

Commercial relationships: none.

Corresponding author: Bilge Sayim.

Email: bilge.sayim@psy.unibe.ch.

Address: Institute of Psychology, University of Bern, Bern, Switzerland.

References

- Andriessen, J. J., & Bouma, H. (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research*, *16*(1), 71–78.
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12):13, 1–18, doi:10.1167/9.12.13. [PubMed] [Article]
- Banks, W. P., Bachrach, K. M., & Larson, D. W. (1977). The asymmetry of lateral interference in visual letter identification. *Perception & Psychophysics*, *22*(3), 232–240.
- Banks, W. P., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception and Psychophysics*, *19*, 361–367.
- Bernard, J.-B., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target-flanker similarity. *Journal of Vision*, *11*(8):1, 1–16, doi:10.1167/11.8.1. [PubMed] [Article]
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*(5241), 177–178.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.
- Chastain, G. (1981). Inhibition of feature extraction with multiple instances of the target feature in different orientations. *Psychological Research*, *43*, 45–56.
- Chastain, G. (1982). Confusability and interference between members of parafoveal letter pairs. *Perception & Psychophysics*, *32*, 576–580.
- Chung, S. T. L., & Legge, G. E. (2009). Precision of position signals for letters. *Vision Research*, *49*(15), 1948–1960.
- Coates, D. R., 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), 1–20, doi: 10.1177/2041669517705447.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.
- Cohen, J. (1988). *Statistical power analysis for the*

- behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Erlbaum.
- Ester, E. F., Klee, D., & Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1022–1033.
- Estes, W. K. (1972). Interactions of signal and background variables in visual processing. *Perception & Psychophysics*, *12*(3), 278–286.
- Estes, W. K., Allmeyer, D. H., & Reder, S. M. (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics*, *19*, 1–15.
- Freeman, J., Chakravarthi, R., & Pelli, D. G. (2012). Substitution and pooling in crowding. *Attention, Perception, & Psychophysics*, *74*, 379–396.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201.
- Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding changes appearance. *Current Biology*, *20*(6), 496–501.
- 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, 1–21, doi:10.1167/14.6.2. [PubMed] [Article]
- Hanus, D., & Vul, E. (2013). Quantifying error distributions in crowding. *Journal of Vision*, *13*(4):17, 1–27, doi:10.1167/13.4.17. [PubMed] [Article]
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*(6598), 334–337.
- Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision*, *15*(6):5, 1–18, doi:10.1167/15.6.5. [PubMed] [Article]
- Huckauf, A. & Heller, D. (2002). What various kinds of errors tell us about lateral masking effects. *Visual Cognition*, *9*(7), 889–910.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*, 171–216.
- Koenderink, J., Valsecchi, M., van Doorn, A., Wagemans, J., & Gegenfurtner, K. (2017). Eidolons: Novel stimuli for vision research. *Journal of Vision*, *17*(2):7, 1–36, doi:10.1167/17.2.7. [PubMed] [Article]
- Korte, W. (1923). Über die Gestaltauffassung im indirekten Sehen [On the apprehension of gestalt in indirect vision]. *Zeitschrift Für Psychologie*, *93*, 17–82.
- Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, *8*(2), 255–279.
- Krumhansl, C. L., & Thomas, E. A. C. (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception & Psychophysics*, *21*, 269–279.
- Levi, D. M. (2008). Crowding—an essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654.
- Levi, D. M., Klein, S. A., & Yap, Y. L. (1987). Positional uncertainty in peripheral and amblyopic vision. *Vision Research*, *27*, 581–597. <
- Livne, T. & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, *7*(2):4, 1–12, doi:10.1167/7.2.4. [PubMed] [Article]
- Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? *Journal of Vision*, *10*(3):1, 1–14, doi:10.1167/10.3.1. [PubMed] [Article]
- Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, *12*(10):13, 1–14, doi:10.1167/12.10.13. [PubMed] [Article]
- Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision*, *13*(13):10, 1–10, doi:10.1167/13.13.10. [Article]
- 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, 58–70, doi:10.1167/5.1.6. [PubMed] [Article]
- Metzger, W. (1936). *Gesetze des Sehens*. [Laws of seeing]. Frankfurt am Main, Germany: Kramer.
- Nandy, A. S., & Tjan, B. S. (2007). The nature of letter crowding as revealed by first-and second-order classification images. *Journal of Vision*, *7*(2):5, 1–26, doi:10.1167/7.2.5. [PubMed] [Article]
- 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.
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Feature integration from detection. *Journal of Vision*, *4*(12):12, 1136–1169, doi:10.1167/4.12.12. [PubMed] [Article]
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of*

- Vision*, 7(2):20, 1–36, doi:10.1167/7.2.20. [PubMed] [Article]
- Petrov, Y. & Popple, A. V. (2007). Crowding is directed to the fovea and preserves only feature contrast. *Journal of Vision*, 7(2):8, 1–9, doi:10.1167/7.2.8. [PubMed] [Article]
- Pöder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, 7(2):23, 1–12, doi:10.1167/7.2.23. [PubMed] [Article]
- Rosenholtz, R., Huang, J., & Ehinger, K. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a Lossy representation in peripheral vision. *Frontiers in Psychology*, 3(13), 1–15.
- Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision*, 9(2):5, 1–11, doi:10.1167/9.2.5. [PubMed] [Article]
- Sayim, B., & Cavanagh, P. (2013). Grouping and crowding affect target appearance over different spatial scales. *PLoS ONE* 8(8): e71188, doi:10.1371/journal.pone.0071188.
- Sayim, B., Myin, E., & van Uytven, T. (2015). Prior knowledge modulates peripheral color appearance. *Proceedings of the International Colour Association (AIC), Tokyo Midterm meeting, 2015*, 1147–1152.
- Sayim, B., Westheimer, G., & Herzog, M. H. (2008). Contrast polarity, chromaticity, and stereoscopic depth modulate contextual interactions in vernier acuity. *Journal of Vision*, 8(8):12, 1–9, doi:10.1167/8.8.12. [PubMed] [Article]
- Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science*, 21(5), 641–644.
- Sayim, B., Westheimer, G., & Herzog, M. H. (2011). Quantifying target conspicuity in contextual modulation by visual search. *Journal of Vision*, 11(1):6, 1–11, doi:10.1167/11.1.6. [PubMed] [Article]
- Strasburger, H. (2005). Unfocused spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, 5(11):8, 1024–1037, doi:10.1167/5.11.8. [PubMed] [Article]
- Strasburger, H. (2014). Dancing letters and ticks that buzz around aimlessly: On the origin of crowding. *Perception*, 43(9), 963–976, doi.org/10.1068/p7726.
- Strasburger, H., Harvey, L. O., Jr., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, 49(6), 495–508.
- Strasburger, H., & Malania, M. (2013). Source confusion is a major cause of crowding. *Journal of Vision*, 13(1):24, 1–20, doi:10.1167/13.1.24. [PubMed] [Article]
- Styles, E. A., & Allport, D. A. (1986). Perceptual integration of identity, location and colour. *Psychological Research*, 48, 189–200.
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, 32(7), 1349–1357.
- Wallace, M. J., Chiu, M. K., Nandy, A. S., & Tjan, B. S. (2013). Crowding during restricted and free viewing. *Vision Research*, 84, 50–59.
- Wang, H., He, X., & Legge, G. E. (2014). Effect of pattern complexity on the visual span for Chinese and alphabet characters. *Journal of Vision*, 14(8):6, 1–17, doi:10.1167/14.8.6. [PubMed] [Article]
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15(4), 160–168.
- Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A*, 14(9), 2057–2068.
- Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, 82, 184–199.
- Wolford, G., & Shum, K. H. (1980). Evidence for feature perturbations. *Perception & Psychophysics*, 27, 409–420.
- Zhang, J.-Y., Zhang, G.-L., Liu, L., & Yu, C. (2012). Whole report uncovers correctly identified but incorrectly placed target information under visual crowding. *Journal of Vision*, 12(7):5, 1–11, doi:10.1167/12.7.5. [PubMed] [Article]
- 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, 44–53.

Appendix A: Error consistency

To test if similar error rates occurred with identical stimuli, we correlated the error rates of each error type in the first and second target drawings of Experiment 1. The correlations were relatively low (average: $r = 0.22$; Omission: $r = 0.31$; Addition: $r = 0.19$; Truncation: $r = 0.21$; Extension: $r = 0.06$; Translation: $r = 0.22$; Rotation: $r = 0.32$; Distortion: $r = 0.20$). Even though there was no difference between the error rates in the first and the second presentation (repeated measures ANOVA, $F(1, 4) = 7.344$, $p = 0.054$, there was a clear trend for better overall performance in the second

presentation (second presentation: average error rate = 0.136, $SE = 0.016$; first presentation: average error rate = 0.156, $SE = 0.010$). This trend was largely due to relatively low Omission error rates in the second (average error rate = 0.16, $SE = 0.025$) compared to the first presentation (average error rate = 0.25, $SE = 0.046$). Hence, training could have played a role in the reduction of the Omission error rate. The other error types did not show any trend for an increase or decrease of performance in the second half of the experiment.

Taken together, the resulting correlations indicate that stimulus appearance varied relatively strongly. Alternatively, the drawing method, with its simultaneous measurement of several error dimensions, and the mutual dependence of error types may have played a role in this result. However, as crowding yields high levels of perceptual indeterminacy, we did not expect very high correlations, and it is unclear what magnitude of correlations can be expected in this paradigm. Future studies are needed to elaborate to what extent the error rates of the different error types vary when the same stimulus is presented multiple times.

Appendix B: Target identification

To investigate how the different error types in Experiment 1 were related to identification performance, we asked four new, naive observers to indicate the identity of the targets in the drawings. Observers were presented with the drawings of the complex

flanker condition at 12° eccentricity and with the baseline drawings. The task was to freely name each target. There were no time constraints, and observers viewed each stimulus until they responded. Responses to the numbers and letters, both rotated and in their cardinal orientation, were classified as “correct” when identifying the original target and “incorrect” when failing to do so (meaningless symbols were excluded from the analysis; however, it is interesting to note that meaningful labels were given in 43%, $SE = 3.2\%$, to the drawings of meaningless targets). In the flanked condition, the average proportion of correctly identified targets was 0.27 ($SE = 0.04$) compared to 0.91 ($SE = 0.03$) in the baseline condition. Hence, identification performance strongly deteriorated when the targets were crowded compared to unflanked targets. Next, we correlated the error rates of each error type with identification performance in the flanked condition. All error rates, except Extension errors, were negatively correlated with identification performance, i.e., high (low) error rates were associated with poor (good) identification performance. In descending order, the correlation coefficients (Pearson’s ρ), were Omission ($\rho = -0.40$), Rotation ($\rho = -0.29$), Truncation ($\rho = -0.27$), Translation ($\rho = -0.25$), Addition ($\rho = -0.13$), Distortion ($\rho = -0.05$), Extension ($\rho = 0.04$). The results indicate that the error types are related to identification performance to different extents. For example, Omission errors are a rather good predictor of performance, whereas Extension errors are not. We suggest that the magnitude of the correlations reflects how much a particular error type renders a target unidentifiable in crowding.