Configural superiority for varying contrast levels

Pieter Moors¹ · Thiago Leiros Costa¹ · Johan Wagemans¹

Published online: 18 November 2019
© The Psychonomic Society, Inc. 2019

Abstract

Observers can search for a target stimulus at a particular speed and accuracy. Adding an identical context to each stimulus can improve performance when the resulting stimuli form clearly discriminable configurations. This search advantage is known as the configural superiority effect (CSE). A recent study showed that embedding these stimuli in noise revealed lower contrast thresholds for part-stimuli compared to configural stimuli (Bratch et al., *Journal of Experimental Psychology: Human Perception and Performance*, 42(9), 1388–1398, 2016). This contrasts with the accuracy advantages traditionally associated with CSEs. In this study, we aimed to replicate the results of Bratch et al. and asked whether the benefit for part-stimuli held across the full psychometric function. Additionally, we tested whether embedding the stimuli in noise was crucial for obtaining their result and whether different contrast definitions affected the results. Furthermore, we used control stimuli that were more directly comparable. Our results showed a detection benefit for the Gestalt context stimuli in all conditions. Together, these results are in line with the literature on CSEs and do not seem to support the recent claim that Gestalts are processed less efficiently than part stimuli. Inspired by this, we sketch how contrast manipulations could be an additional tool to study how Gestalts are processed.

Keywords Perceptual organization · Visual perception · Visual search

Introduction

Configural Gestalts are characterized by part-whole relationships (i.e., stimuli that are qualitatively different from the sum of their parts) and are central to human visual perception. The study of these phenomena (or, perceptual organization more broadly) has frequently relied on phenomenology alone through showcasing the effects (reviewed by Wagemans et al., 2012 and Wagemans, 2018). In the 1970s the information processing literature offered a few examples of quantification of these phenomena (Kolers & Pomerantz, 1971; Pomerantz & Garner, 1973). However, more sophisticated psychophysical approaches remained absent for a long time. In the 1990s, this started to change with the quantitative study of grouping by proximity (Kubovy & Wagemans, 1995) and contour integration (Field, Hayes, & Hess, 1993; Kovacs & Julesz, 1993). As a result, psychophysical methods are now

part of the standard approach to study mid-level vision (e.g. (Claessens & Wagemans, 2008; Machilsen, Pauwels, & Wagemans, 2009; Machilsen & Wagemans, 2011; Panis, De Winter, Vandekerckhove, & Wagemans, 2008; Poljac, de-Wit, & Wagemans, 2012). The main conclusion from this literature seems to be that Gestalt-like grouping phenomena generally have a beneficial effect on performance in a psychophysical task. In these studies, the stimulus intensity variable of interest is usually unrelated to physical stimulus intensity as such, and often defined as the degree of contour alignment, for example. In comparison, in low-level vision studies stimulus intensity is nearly always defined in terms of stimulus contrast. An interesting set of exceptions can be found in the recent work from Jason Gold and colleagues (Gold, 2014a, 2014b; Gold et al., 2014). In these studies, the goal was to assess the processing efficiency in phenomena related to perceptual organization such as (a)modal completion, size contrast, configural superiority, or face processing. The general observation from all these studies was that there seems to be a cost (i.e., an inefficiency) of information processing related to Gestalt effects.

The present study was motivated by the results reported by (Bratch et al., 2016). Here, the authors used the configural superiority effect (CSE) with the classic triangle and arrow displays (Pomerantz, Sager, & Stoever, 1977). In the CSE, detection of a target element among a set of distractors is facilitated by adding an identical context stimulus to the

Pieter Moors and Thiago Leiros Costa contributed equally to this work.

✉ Pieter Moors
pieter.moors@kuleuven.be

¹ Laboratory of Experimental Psychology, Department of Brain and Cognition, KU Leuven, Tiensestraat 102 - box 3711, BE-3000 Leuven, Belgium

elements (see Fig. 1). This effect is somewhat counterintuitive from a straight bottom-up feature integration account of vision, as the identical context should mostly increase processing load (e.g., due to lateral masking), and make the stimuli more similar. However, when the interaction between the target and uninformative context leads to emergent features with Gestalt-like qualities, visual search is faster and more accurate (for a review, see Pomerantz & Cragin, 2015). Interestingly, stimuli that induce a CSE have been shown to engage V1 and the lateral occipital cortex in different ways when compared to control stimuli (or their constituent parts; Kubilius, Wagemans, & Op de Beeck, 2011; Fox, Harel, & Bennett, 2017; Kubilius et al., 2011). That is, LOC seems to be involved in processing the configural stimuli, while V1 is more concerned with the processing of the local elements. This is supported by a case study in a patient with bilateral LOC lesions who showed an advantage for processing parts rather than wholes (de Wit, Kubilius, Op de Beeck, & Wagemans, 2013). Thus, the CSE seems to be associated with the representations of configurations in upstream areas of the occipito-temporal cortex rather than the representations of isolated stimulus features in downstream areas of extrastriate cortex.

To assess the information processing behind the CSE, Bratch et al. used a response classification method together with threshold estimation. The goal was to assess processing

efficiency of the configural stimuli by measuring the threshold ratio between human and ideal observers for these stimuli. The authors embedded the stimuli in noise to be able to derive classification images, which could provide insight into which parts of the stimuli were used to generate a response. Interestingly, Bratch et al. observed a classical CSE in response times, but contrast thresholds were lower for the part-stimuli compared to the Gestalt context stimuli. The authors claimed that the stimuli that induced a CSE were associated with lower processing efficiency. In a second experiment, Bratch et al. used a control stimulus consisting of the same local parts combined in a way that does not lead to a CSE or even to a configural inferiority effect (as first reported in Pomerantz et al., 1977, see Fig. 1). As expected, the authors now found no response time difference between conditions. However, they again observed lower contrast thresholds for part-stimuli compared to the control context stimulus. The authors interpreted this latter difference as a consequence of an increase in internal noise due to the context stimulus.

The results reported in Bratch et al. are a potentially interesting anomaly in a literature where detection benefits for Gestalts are consistently documented. The goal of the current study was to further probe the effects observed in Bratch et al. (2016) and to better understand the effect of contrast manipulations on the processing of Gestalt stimuli. Our main point of departure is the fact that Bratch et al. measured contrast thresholds adaptively with a staircase method. Adaptive methods allow to efficiently characterize an observer's threshold, usually without consideration of the slope of the psychometric function. However, in this particular case, measuring the entire psychometric function could reveal more precisely how these stimuli are processed along the stimulus contrast dimension. That is, it might be the case that part-stimuli show lower contrast thresholds compared to Gestalt context stimuli, but that the opposite is true for the slopes of psychometric function. Such a situation might require a more complex interpretation where part-stimuli initially show a detection benefit, yet as soon as the Gestalt cues are sufficiently processed, Gestalt stimuli show a detection benefit. Thus, our first goal was to use the classic CSE stimuli (as in Bratch et al.) together with the method of constant stimuli to estimate psychometric functions for configural Gestalt stimuli and their non-Gestalt controls. Second, Bratch et al. embedded the stimuli in white noise. This embedding was crucial for their response classification method, yet we reasoned it could also critically change the way in which these Gestalt stimuli were processed. Therefore, we presented these stimuli either at very low stimulus intensities against a homogeneous background or we embedded them in noise. Third, Bratch et al. performed two separate experiments with different types of context being added to the baseline parts, leading to comparison of stimuli with different numbers of pixels and also different overall display sizes (i.e., a Gestalt context in Experiment 1 and a

Fig. 1 Examples of displays used to induce the configural superiority effect (CSE). The same base display (top) coupled with identical contexts (middle) will induce a CSE (bottom left) when this combination leads to emergent features (such as closure and intersections). When no such emergent features are present (right), no CSE will be found and discrimination performance will possibly be worse compared to the base displays (i.e., configural inferiority effect). In the rest of the paper, we refer to the base stimuli as “part” or “no context,” to the CSE displays as “Gestalt context” and to the CSI displays as “bad context” or “control context”

control context in Experiment 2; see Fig. 1). As a result, the critical comparison between these conditions was suboptimal and involved a comparison between stimuli containing different numbers of pixels (i.e., part vs. Gestalt or part vs. control) and different display sizes. In principle, it could be the case that this difference is driving the observed difference in contrast thresholds. Therefore, we decided to include the Gestalt and control context together with the part stimuli in a single experiment, with the same display size and same amount of pixels (for the context stimuli), allowing us to compare all stimuli. Furthermore, Bratch et al. used a variant of the root mean squared (RMS) contrast to quantify stimulus contrast. The number of pixels in the stimulus display affects the range of contrast values one can obtain for the same stimulus intensity in terms of percentage deviation from the background luminance (Weber contrast). Throughout this paper, we mostly rely on Weber contrast, but we included an explicit comparison with RMS contrast as well.

In sum, the three specific goals formulated above can be translated into three specific questions:

1. How does the processing of these stimuli (Gestalt) vary along the stimulus contrast dimension?
2. How does the addition of external noise influence the processing of these stimuli? Are the results fundamentally different depending on the presence of external noise?
3. Are these Gestalt-like phenomena still associated with lower contrast thresholds when the control stimuli share exactly the same local parts and are matched for size? Might part of the effects be explained by contrast definitions that are sensitive to the number of pixels?

A reanalysis of the data of Bratch et al. would allow us to already partially answer our first question. Although the data were collected using an adaptive staircase method, it is possible to estimate the average psychometric function across all observers relying on the aggregated data. This estimate is far from optimal, but it might already provide some intuition on how the stimuli are processed along the contrast range. Below, we report the results of this reanalysis (which, importantly, mimic the observations from our experiment). Subsequently, we report the data we collected in a set of 20 new observers.

Reanalysis of the data reported by Bratch et al.

We obtained the data from Bratch et al. with the goal to get a first impression of how stimulus processing evolved along the stimulus contrast dimension. Here, we briefly summarize the main conclusions, as depicted in Fig. 2. The contrast units used here were the same as in the original work (i.e., RMS contrast). The top panels show the mean response time data from both

experiments. It is clear that a CSE was observed in Experiment 1, but no longer in Experiment 2 (in line with Pomerantz et al., 1977). The middle panels show the predicted psychometric functions that were estimated from the aggregated data. Interestingly, in Experiment 1 the detection benefit for the part stimuli is apparent in the low contrast range, whereas the psychometric function for the Gestalt stimuli crosses the one for the part stimuli at a particular level of stimulus contrast. In line with our earlier reasoning, not only the thresholds, but also the slopes of both psychometric functions differ. This contrasts with the psychometric functions obtained in Experiment 2, where no CSE was observed in response times. Here, the psychometric functions clearly dissociate, and a detection benefit for part stimuli is observed across the full contrast range, without an apparent difference between slopes. The qualitative difference between the psychometric functions in Experiments 1 and 2 already indicates a different type of processing for the Gestalt and control stimuli, despite the similarity in difference in contrast thresholds reported in the paper. The results of the ideal observers are depicted in the bottom panels. As expected, the ideal observers are not influenced at all by the (non-)configural nature of the Gestalt or control stimulus. Indeed, the psychometric functions for both experiments are nearly identical. This makes sense because the ideal observer is not sensitive to the particular configuration embedded in the noise pattern (it only “searches for” the odd line stimulus). Nevertheless, it reveals a critical difference between human and ideal observers.

In sum, this reanalysis indicates a data pattern that would be expected based on the reasoning that configural processing is disrupted at lower contrast levels, but shows a clear benefit compared to part stimuli when it starts developing at a particular stimulus contrast. Encouraged by this reanalysis, we set out to replicate these results more formally relying on a design that (1) allowed us to estimate the full psychometric function rather than a single threshold, (2) allowed the comparison between the presence and absence of external noise and (3) allowed for a direct comparison between the three types of stimuli used.

Methods

Participants

Twenty participants took part in the experiment. All had normal or corrected-to-normal vision. They were all naive with respect to the goal of the study, and provided signed informed consent prior to starting. This study was approved by the local ethics committee of the university (SMEC). The experiment was divided in two sessions, each of which lasted approximately 1 h, and participants completed both sessions on separate days (most of the time separated by 24 h).

Fig. 2 Reanalysis of Bratch et al. (2016). Stimulus contrast is expressed in root mean squared (RMS) contrast. **(A)** Average response time data for Experiment 1. Gestalt context stimuli elicit faster responses. **(C)** Accuracy data averaged across observers, and predicted psychometric functions (without taking individual differences into account). The psychometric functions intersect at about 75% correct, suggesting a part benefit at low stimulus intensities and a Gestalt benefit at higher ones. **(E)** Predicted psychometric functions for the ideal observer. For the lower intensity range, the same effect is observed, yet the part benefit remains

across the full range of the psychometric function. **(B)** Average response time data for Experiment 2. Control context stimuli elicit no difference in response times. **(D)** Accuracy data across observers, and predicted psychometric functions (without taking individual differences into account). In contrast to **(C)**, the psychometric functions no longer intersect, suggesting a part benefit across the full stimulus intensity range. **(F)** Predicted psychometric functions for the ideal observer. A part benefit is observed across the full intensity range. The code for this figure can be found at <https://osf.io/zcsd8/>

Apparatus

All stimuli were displayed on a linearized CRT monitor (Sony GDM F520) with a refresh rate of 60 Hz, a diagonal length of

50 cm, and a resolution of 1,024 x 768 pixels. Participants were positioned in a dark room at a viewing distance of 57 cm from the monitor by means of a chin rest. Stimulus presentation and response registration was controlled by

custom software programmed in Python 2.7, mainly relying on the PsychoPy library (Peirce et al., 2019). The Bits# system (mono++ mode) was used to manipulate stimulus contrast at a 14-bit resolution. The background of the monitor was set to mid-gray (38 cd/m²).

Stimuli

We used very similar stimuli to the ones used in Bratch et al. (2016), who modeled their stimuli after Pomerantz et al. (1977) (see Fig. 3). The part stimulus consisted of a line rotated 45° counterclockwise relative to vertical in the odd quadrant, whilst the other quadrants were filled with lines oriented 45° clockwise relative to vertical (28 x 3 pixels, except for the first and last part of the line which consisted of 2 pixels). For the Gestalt context condition, two abutting lines (30 x 2 pixels) were added which formed a right angle. This context produced a triangle in the odd quadrants, and an arrow shape in the other quadrants. In the bad context condition, the same abutting lines were added, but now in such a way it did not generate the percept of closure as in the Gestalt context stimuli. The total size of each stimulus (all quadrants) was 160 x 160 pixels. The no context line stimuli always appeared on the same location (barring which quadrant was defined as odd) and depending on the context condition, particular context stimuli were added to these lines. Contrast of the pixels (c_i) at pixel location i was defined as:

$$c_i = \frac{l_i - L}{L}$$

where l_i is the luminance of the i th pixel and L is the average background luminance. In practice, contrast was varied before external noise was added to the stimulus. This is similar to how Bratch et al. defined stimulus contrast for each pixel (i.e., Weber contrast), but they converted these values to RMS contrast to be used in the staircase procedure. In the noise experiment, external noise was added by randomly selecting pixel intensities from a Gaussian distribution centered on 0 with a variance of 0.0625. After adding noise, contrast values were clipped at ± 2 standard deviations to avoid exceeding the minimal and maximal luminance values that could be presented on the monitor.

Procedure and design

The noise and no-noise conditions were run in separate sessions, and the order of completion of these conditions was counterbalanced across participants (odd participant numbers starting with noise, and even with no-noise). All trial types (part, Gestalt context, and control context) of all contrast levels were randomly intermixed. Each trial started with the presentation of a fixation dot (500 ms), after which the display was presented for 300 ms. After stimulus offset, participants had to indicate the location of the odd quadrant as fast and accurately as possible, using the numerical keyboard (1 = bottom-left; 2 = bottom-right; 4 = top-left; 5 = top-right). After each response, there was a 500-ms intertrial interval (blank screen) before proceeding to the next trial. Participants could take a break after every 100 trials. In the noise and no-noise conditions, seven different (fixed) stimulus

Fig. 3 Displays used in the experiments. Columns indicate stimuli with the same contrast. Rows indicate the same stimulus class (part, Gestalt context, control context). Contrast levels have been adjusted for visibility and are not depicted as in the experiment

contrast levels were used. These were logarithmically spaced between the minimal and maximal contrast value used. The range of values was determined based on pilot experiments in order to ensure that the lowest value resulted in chance performance, and the highest value in near-perfect performance. For the noise condition, the minimum and maximum were 0.1 and 0.3. For the no-noise experiment, we used 0.01 to 0.08. For each combination of stimulus type and stimulus contrast, we ran 60 trials. This resulted in a total of 1,260 trials for each session, or 2,520 trials in total for each participant.

Data analysis

We used R 3.5.1 and RStudio (Team & Others, 2015) for all our analyses (for a complete overview of the environment used, see the R Markdown document documenting all code used to analyze the data). All code and data can be found on <https://osf.io/w3ueb/>.

Psychometric functions Psychometric functions were analyzed using a Bayesian generalized linear mixed modelling framework (Bürkner & Others, 2017; Moscatelli, Mezzetti, & Lacquaniti, 2012; Yssaad-Fesselier & Knoblauch, 2006), relying on the R package *brms* (Bürkner & Others, 2017). Although fitting single psychometric functions for each observer-condition combination would have been definitely feasible, the current approach has a lot of advantages. That is, modelling all data simultaneously using a hierarchical approach can provide more statistical power compared to the

two-step approach of estimating parameters of the psychometric function and statistically analyzing afterwards. Furthermore, it is straightforward to flexibly adjust the guessing parameter as well as the lapse parameter in the Bayesian framework (or even put predictors on them if one wishes to do so). Lastly, it is still possible to include inter-individual variability in the statistical model, yet the estimates for the individuals are influenced by other individuals (due to the hierarchical nature of the model), yielding better estimates compared to the two-step approach. We used a logit link function to model the sigmoidal pattern of the psychometric function. Because we wanted to be able to compare differences between the noise and no-noise conditions, we standardized both intensity values to the same mean and standard deviation (z-transform). The lapse rate of the psychometric function was fixed at .05. We initially tried to model individual variability in the lapse rates of the psychometric function, yet this resulted in poor convergence of the MCMC chains. The guess rate was fixed to .25, as the task was 4-AFC. *brms* is a software package providing a convenient front-end to STAN where complex Bayesian models can be fit using Hamiltonian MCMC methods (Carpenter et al., 2017). We put a Student-t prior on all regression weights, and a Cauchy prior on the parameters capturing the random effects (as in (Wallis et al., 2017)). All random effects were included, yet uncorrelated to ensure model convergence. The fitted model was the following:

$$\text{response} = .25 + (1 - .25 - .05) * \text{logit} \left(\text{stimulus intensity} * \text{stimulus type} * \text{condition} + \left(\text{stimulus intensity} * \text{stimulus type} * \text{condition} \parallel \text{participant} \right) \right)$$

All categorical variables were dummy-coded. As stimulus intensity differed substantially between the noise and no-noise conditions, we standardized (z-score) both intensity variables separately, such that their coefficients were comparable. As outlined in (Knoblauch & Maloney, 2012), the location and scale of the link function can be derived as follows:

$$\begin{aligned} \text{location} &= -(\text{intercept}/\text{slope}) \\ \text{scale} &= 1/\text{slope} \end{aligned}$$

Given our guess and lapse rate, the location equals the 60% threshold, and the scale quantifies the variability around this threshold. In the *Results* section, we report unstandardized estimates for comparisons within noise and no noise conditions. When comparing across these conditions, we kept the values on the standardized scale. Inference on these quantities was performed by summarizing the posterior distributions of various pairwise comparisons. Rather than applying an arbitrary threshold to claim the presence or absence of an “effect”, we quantify our (un)certainly of differences between

conditions by calculating the posterior probability that thresholds are smaller in one condition versus the other (see *Results*).

We ran four chains with 4,000 iterations each (2,000 of each were considered warm-up). The *adapt_delta* parameter was set to .8 and the *max_treedepth* to 10. Running this model yielded MCMC chains that were converging, according to diagnostics such as effective sample size, the R-hat statistic, and the absence of divergent transitions.

Response times We used a linear mixed-effects model with a lognormal distribution to model response times. The lognormal is a skewed distribution that allows one to model and capture positive skew that is common in response time (RT) data. We again relied on *brms* to estimate the parameters of this model. The structure of the model was exactly the same as for the psychometric function, except for the response distribution now being lognormal rather than binomial. Although it is possible to include predictors for the standard deviation of the distribution, we limited ourselves to modelling the mean

response times only. We trimmed the RT data at response times of 3 s, which we considered to be unreasonably long for the current task. On average, this resulted in 1.45% of the data per participant being removed with a maximum of 8%.

We ran four chains with 2,000 iterations each (1,000 of each were considered warm-up). The *adapt_delta* argument was set to .8 and the *max_treedepth* to 10. Running this model yielded MCMC chains that were converging, according to diagnostics such as effective sample size, the R-hat statistic, and the absence of divergent transitions.

Results

Psychometric functions

Figure 4 depicts the behavioral data, overlaid with the psychometric functions estimated by the model. It is clear that, for both stimulus types, the same ordering of psychometric functions was observed. As soon as performance is above chance, the Gestalt context stimulus yielded better performance compared to the part stimulus, and the control context. Furthermore, the Gestalt context stimulus always yielded lowest 60% contrast thresholds compared to the part stimulus, and the control context. For both stimulus types, the control context yielded the highest contrast threshold, and the part stimulus was always in-between.

Each scatterplot in Fig. 5 depicts two-dimensional posterior distributions for pairwise comparisons between thresholds and slopes for all conditions. As can be derived from the legend in each panel, all posterior probabilities suggest that thresholds for the Gestalt context are lower than for the part stimulus, and the control context (in both noise and no noise conditions). In addition, part stimuli have lower thresholds compared to the control context. A similar result is obtained for the slopes across the noise and no noise condition, Gestalt contexts elicit steeper slopes compared to the other conditions,

and the same holds for the part stimulus compared to the control context.

An interesting pattern of results emerges when comparing the thresholds and slopes between the noise and no noise conditions within each stimulus condition (Fig. 6). For neither of the comparisons, the thresholds seem to be strongly different. However, for the slopes, it seems to be the case that adding noise to the stimuli yields shallower psychometric functions compared to the no noise stimuli. In a sense, the absence of a difference in threshold is difficult to interpret, as the stimuli can only be compared on a standardized scale. Presumably, this indicates that the stimulus range was well-calibrated, and the midpoint of the intensities yielded similar performance.

Response times

In Fig. 7, the mean response times are depicted together with the model predictions (see Table 1 for a summary of the parameter estimates).

The RTs clearly mimic the accuracy data. Response times increase with stimulus contrast for the parts stimulus and the control context, whereas they decrease for the Gestalt context (i.e., there is a CSE). Stimulus type only shows specific interactions. That is, the control context stimulus differs less from the parts stimulus for noise stimuli, and the slopes associated with stimulus contrast also become more shallow in the noise experiment, for both the Gestalt context as well as the control context. This is clearly visible in the right panel of Fig. 7 where the same average trends are observed, yet the change in response times across stimulus contrast is less strong compared to the no noise stimuli.

Combined analysis

In the final analysis, we were interested to see how the modeled accuracy and RT data were linked with each other.

Fig. 4 Aggregate behavioral data and estimated psychometric functions. Points are mean accuracies, error bars (not always visible) are mean accuracies ± 2 SEM. Lines are the estimated psychometric functions,

and shaded regions denote 95% credible intervals around the psychometric functions

Fig. 5 Posterior distributions for pairwise comparisons between conditions within stimulus type. Each scatter plot depicts three different posterior distributions of thresholds (left) and slopes (right) for pairwise comparisons between all conditions within each stimulus type (no noise

(upper) or noise (lower)). The legend indicates the posterior probability that thresholds or slopes in one condition are smaller than in the condition it is being compared with

To address this, we extracted the estimated accuracies and response times for each iteration in the MCMC procedure, for all unique conditions. We then calculated the differences between accuracies and response times between conditions (e.g., part vs. Gestalt context), respectively, and plot these differences in the scatterplots below. The primary question of interest here is to see whether the RT differences emerge earlier or later in function of increasing stimulus contrast compared to the accuracy data. Figure 8 depicts this separately for the noise and no-noise stimuli, where each pairwise

comparison between all conditions was computed. For both stimulus types, the pattern is qualitatively similar and the effects on both measures are coherently related. As soon as the effects on the accuracy scale emerge, the same happens for response times. This is less clear in Fig. 9. Here, for each condition, the difference between noise and no noise stimulus is depicted. For part and Gestalt context stimuli, the response time effect is largely constant, whereas it seems to evolve for the control context stimuli. The accuracy pattern is very similar across conditions. Interestingly, noise in the stimuli seems

Fig. 6 Posterior distributions for pairwise comparisons between stimulus types within each condition. Each scatter plot depicts three different posterior distributions of thresholds (left) and slopes (right) for pairwise comparisons between stimulus type (no noise vs. noise) within each

condition (part, Gestalt, control). The legend indicates the posterior probability that thresholds or slopes for one stimulus type (e.g., no noise) are smaller compared to the other stimulus type (e.g., noise), for each condition separately

Fig. 7 Aggregate response time data and model predictions. Points are mean response times, error bars (not always visible) are mean response times ± 2 SEM. Lines are the estimated regression lines, and shaded regions denote 95% credible intervals around the regression lines

to facilitate stimulus detection for very low stimulus intensities whereas the opposite occurs at higher intensities. As highlighted earlier, it is difficult to interpret the differences between noise conditions because they are coded on different parts of the stimulus contrast scale.

Discussion

In this study, we asked whether configural superiority effects are still observed when stimuli are presented under very low levels of stimulus intensity or embedded in noise. We observed that Gestalt context stimuli require less contrast to be discriminated as accurately as part stimuli. In addition, control context stimuli elicited an inferiority effect in that they had to be presented at higher contrasts to be discriminated as accurately as the Gestalt context or part stimuli. This pattern of

results was reflected in the response time data, where superiority and inferiority effects were observed as well. The location of contrast detection thresholds was very similar for the noise and no-noise conditions, but the steepness of the psychometric functions did vary considerably between both conditions. That is, detection of no-noise stimuli was associated with steeper slopes compared to detection of stimuli embedded in white noise.

At a first glance, our results suggest a pattern opposite to what is reported by Bratch et al. In fact, our results are fully consistent with theirs, and perfectly replicate their results. The main reason is that, throughout this study, we quantified stimulus contrast as a percentage (i.e., how much it differed from the uniform background). Bratch et al. employed the same approach to construct the stimulus (with some small differences we have discussed in the *Methods* section), yet they quantified stimulus contrast as root mean squared (RMS)

Table 1 Summary of the model parameter estimates for the response time data

Predictors	Estimates	HDI (95%)
Intercept (parts baseline)	-0.68	-0.81 ; -0.54
Stimulus contrast	0.13	0.04 ; 0.21
Gestalt context	-0.20	-0.24 ; -0.15
Control context	0.16	0.11 ; 0.22
Noise experiment	0.10	-0.06 ; 0.26
Stimulus contrast * Gestalt context	-0.25	-0.30 ; -0.20
Stimulus contrast * Control context	0.08	0.04 ; 0.12
Stimulus contrast * Noise experiment	-0.04	-0.11 ; 0.02
Gestalt context * Noise experiment	0.03	-0.01 ; 0.07
Control context * Noise experiment	-0.07	-0.12 ; -0.03
Stimulus contrast * Gestalt context * Noise experiment	0.06	0.02 ; 0.09
Stimulus contrast * Control context * Noise experiment	-0.04	-0.08 ; -0.01

Note. The parts stimulus served as baseline in the regression model. Each estimate should be interpreted relative to the intercept or to the estimated slope of stimulus contrast

Fig. 8 Comparison of the response time and accuracy differences for the different conditions, visualized separately for noise and no noise stimuli. The response time and accuracy effects both refer to *estimated effects*. These were computed by calculating the depicted pairwise differences at

each iteration during the MCMC procedure. One can thus interpret these as posterior distributions of the pairwise differences. Stimulus contrast is visualized by the intensity of the dots and lines (low to high being low to high stimulus contrast). The error bars are 95% credible intervals

contrast. RMS contrast is a measure that is often used for quantifying global image contrast, as it essentially is a measure of variation in pixel intensities. This is where the crucial difference between both measures emerges. A percentage measure is straightforward to calculate when an image consists of two intensity values (as in our case), and hence it is insensitive to the number of pixels a stimulus is comprised of. In comparison, RMS contrast is sensitive to the number of pixels a stimulus contains, as it induces more variation in the global image. Therefore, the contrast range for the part stimuli was different from the context stimuli in Bratch et al., despite the fact that, expressed in percentages, the contrast range was the same. Below we visualize our data in an alternative way to explicitly compare percentage contrast with RMS contrast. It is clear that the qualitative pattern of results for Gestalt and control context stimuli remains the same. Although the

contrast values relate nonlinearly to each other, the mapping is exactly the same for both stimuli. In contrast, for the part stimulus, the RMS contrast measure yields lower values. This implies that, for similar performance, lower thresholds will be obtained for the part stimulus compared to the other two stimulus types, even if the psychometric function of the part stimulus would be exactly the same as for one of the context stimuli. This could lead to the impression that part stimuli lead to lower thresholds than Gestalt context and this could be interpreted as better detection of parts at threshold.

Interestingly, this contrast coding measure also explains why they did not observe a configural inferiority effect. Indeed, when expressed in RMS contrast, RTs are distributed in such a way that an adaptive staircase procedure would yield no difference between the part stimuli and the control context (see Fig. 11).

Fig. 9 Comparison of the response time and accuracy differences between the noise and no noise stimuli, for each condition separately. The response time and accuracy effects both refer to *estimated effects*. These were computed by calculating the depicted pairwise differences at

each iteration during the MCMC procedure. One can thus interpret these as posterior distributions of the pairwise differences. Stimulus contrast is visualized by the intensity of the dots and lines (low to high being low to high stimulus contrast). The error bars are 95% credible intervals

Fig. 10 Comparison between stimulus contrast expressed in percentages or root mean square contrast. The size of the dots indicates the accuracy associated with the data point. For both stimulus presentation conditions, the Gestalt and control context show exactly the same relationship

Despite the consequences of using RMS contrast, one could argue that it is a meaningful measure to use because it makes sense to correct for the number of pixels a stimulus is comprised of. However, we believe the way we quantified stimulus contrast in this study best reflects how the visual system deals with processing these stimuli. That is, in the no-noise condition it would be natural to opt for a percentage measure of stimulus contrast, and the results are qualitatively the same as in the noise condition. Even so, when equating the number of pixels (i.e., comparing the Gestalt and control context stimuli), it is clear that the Gestalt context stimulus always yielded better performance compared to the control context stimulus. In sum, we believe both our results, as well as those reported in Bratch et al., reflect that Gestalt context stimuli

between percentage contrast and RMS contrast. The part stimuli are associated with consistently lower RMS contrast levels, however, yielding lower thresholds for RMS contrasts

yield lower detection thresholds compared to part stimuli and control context stimuli.

Reflecting back on our introduction, these results support the general trend in the literature that good configurations are easier and quicker to detect compared to configurations that do not contain Gestalt properties, or very simple stimuli, such as an array of line stimuli. Although consistent with the literature, our results are, in a sense, maybe counterintuitive. That is, in order to process emergent features, the whole rather than a part of the stimulus needs to be perceived. When a stimulus is presented in an impoverished fashion as in the current study, one would expect that observers fail to perceive some parts of the stimulus and hence a part benefit would be expected. This does not seem to be the case, however, making configural superiority as observed here maybe even more striking compared to traditional studies.

We see great potential in the paradigm introduced by Bratch et al. in the sense that Gestalt psychology has mostly focused on phenomenological demonstrations, rather than psychophysical experiments to provide empirical support for their theoretical premises. This is not to say that Gestalt psychology has not entered the realm of psychophysics yet (Denisova, Feldman, Su, & Singh, 2016; Erlikhman & Kellman, 2016; Spehar & Halim, 2016). Interestingly, most of these studies present stimuli at suprathreshold intensities and disrupt stimulus processing through other manipulations. In this way, we believe that a manipulation of stimulus intensity in terms of stimulus contrast provides a potentially promising new line of research in which processing characteristics of mid-level stimuli can be objectively studied. Indeed, such an approach would allow us to exploit basic paradigms used in low-level vision that have been used to elucidate processing

Fig. 11 Response time data for the noise condition, in function of root mean squared (RMS) contrast. The part stimuli are squashed towards the lower end of the contrast values, rendering them indistinguishable from the control context stimulus. This could explain why Bratch et al. did not observe a consistent configural inferiority effect in their Experiment 2

characteristics. For example, Brown, Breitmeyer, Hale, and Plummer (2018) proposed to use the linearity of contrast response functions to assess whether visual illusions are processed rather in a percept-dependent than in a stimulus-dependent way. The authors assessed illusion magnitudes across different contrast levels and found that the Poggendorf and Ponzo illusions showed a much more nonlinear increase in illusion magnitude compared to the simultaneous brightness illusion, which is frequently explained by a low-level mechanism, based on lateral inhibition. Although the approach of assessing the nonlinearity of the contrast response functions is not directly transferable to our study, it does show the potential for application in studying mid-level visual processing.

Apart from studying Gestalt processing through contrast manipulations, this type of manipulation could also be thought of as a quantification of the “Gestaltness” of a stimulus. Indeed, quantitative assessments of mid-level visual processing have always been quite challenging, and only a few attempts have been successful in providing a quantification of the strength of perceptual grouping. For example, grouping by proximity has been quantified through the pure distance law (Kubovy & Wagemans, 1995) or Garner interference (Pomerantz & Schwaizberg, 1975). Wei, Zhou, and Chen (2018) proposed to quantify perceptual grouping in general by measuring tilt aftereffects. Last, Pomerantz and Portillo (2011) suggested to use configural superiority itself as an index of grouping strength. Here we propose that the way in which a stimulus responds to stimulus contrast manipulations might be a fruitful approach to quantify the strength of perceptual grouping. Indeed, the stimuli used in this study relied on a particular topological distinction (closed vs. open) (Chen, 2005) and might have generated a strong difference signal. Other stimuli that contain other types of emergent features might be less salient, and the configural superiority elicited by such stimuli might be less resistant to contrast and saliency manipulations as what has been presented here. At this point, this is just an expression of an intuition about a promising avenue. Obviously, more studies are needed to see whether this actually turns out to be correct.

Conclusion

In this study we asked whether configural superiority is observed for stimuli presented at low stimulus intensities. We measured contrast detection thresholds for stimuli varying in the degree of configurality and observed that stimuli eliciting a Gestalt were discriminated at lower contrast levels compared to their parts, or a control context not traditionally eliciting configural superiority. This configural superiority effect was also observed in response times, and the qualitative pattern of results did not change depending on whether stimuli were

embedded in white noise or not. We conclude that Gestalt phenomena are not restricted to high stimulus intensities only, and propose that contrast manipulations could provide a useful tool to assess how different Gestalts are processed.

Acknowledgements PM is supported by the Research Foundation – Flanders through a postdoctoral fellowship. TLS is supported by a Marie Skłodowska Curie fellowship. JW is supported by the Research Foundation – Flanders through Methusalem Funding (METH/14/02).

Open practices statement All code of the experiments and data analysis, as well as the data can be found at <https://osf.io/w3ueb/>. No experiments in this study were formally preregistered.

References

- Bratch, A., Barr, S., Bromfield, W. D., Srinath, A., Zhang, J., & Gold, J. M. (2016). The impact of configural superiority on the processing of spatial information. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9), 1388–1398.
- Brown, J. M., Breitmeyer, B. G., Hale, R. G., & Plummer, R. W. (2018). Contrast sensitivity indicates processing level of visual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, 44(10), 1557–1566.
- Bürkner, P.-C., & Others. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1). Retrieved from <https://www.osti.gov/biblio/1430202>
- Chen, L. (2005). The topological approach to perceptual organization. *Visual Cognition*, 12, 553–637.
- Claessens, P. M. E., & Wagemans, J. (2008). A Bayesian framework for cue integration in multistable grouping: Proximity, collinearity, and orientation priors in zigzag lattices. *Journal of Vision*, 8(7), 33.1–23.
- Denisova, K., Feldman, J., Su, X., & Singh, M. (2016). Investigating shape representation using sensitivity to part- and axis-based transformations. *Vision Research*, 126, 347–361.
- de Wit, L. H., Kubilius, J., Op de Beeck, H. P., & Wagemans, J. (2013). Configural Gestalts remain nothing more than the sum of their parts in visual agnosia. *I-Perception*, 4(8), 493–497.
- Erlikhman, G., & Kellman, P. J. (2016). Modeling spatiotemporal boundary formation. *Vision Research*, 126, 131–142.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local “association field.” *Vision Research*, 33(2), 173–193.
- Fox, O. M., Harel, A., & Bennett, K. B. (2017). How Configural Is the Configural Superiority Effect? A Neuroimaging Investigation of Emergent Features in Visual Cortex. *Frontiers in Psychology*, 8, 32.
- Gold, J. M. (2014a). A Perceptually Completed Whole Is Less Than the Sum of Its Parts. *Psychological Science*, 25(6), 1206–1217.
- Gold, J. M. (2014b). Information processing correlates of a size-contrast illusion. *Frontiers in Psychology*, 5, 142.
- Gold, J. M., Barker, J. D., Barr, S., Bittner, J. L., Bratch, A., Bromfield, W. D., ... Srinath, A. (2014). The perception of a familiar face is no more than the sum of its parts. *Psychonomic Bulletin & Review*, 21(6), 1465–1472.
- Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R*. <https://doi.org/10.1007/978-1-4614-4475-6>
- Kolers, P. A., & Pomerantz, J. R. (1971). Figural change in apparent motion. *Journal of Experimental Psychology*, 87(1), 99–108.

- Kovacs, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences*, 90(16), 7495–7497.
- Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2011). Emergence of perceptual Gestalts in the human visual cortex: the case of the configural-superiority effect. *Psychological Science*, 22(10), 1296–1303.
- Kubovy, M., & Wagemans, J. (1995). Grouping by Proximity and Multistability in Dot Lattices: A Quantitative Gestalt Theory. *Psychological Science*, 6(4), 225–234.
- Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision*, 9(12), 11.1–11.
- Machilsen, B., & Wagemans, J. (2011). Integration of contour and surface information in shape detection. *Vision Research*, 51(1), 179–186.
- Moscattelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, Vol. 12, pp. 26. <https://doi.org/10.1167/12.11.26>
- Panis, S., De Winter, J., Vandekerckhove, J., & Wagemans, J. (2008). Identification of everyday objects on the basis of fragmented outline versions. *Perception*, 37(2), 271–289.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203.
- Poljac, E., de Wit, L., & Wagemans, J. (2012). Perceptual wholes can reduce the conscious accessibility of their parts. *Cognition*, 123(2), 308–312.
- Pomerantz, J. R., & Cragin, A. I. (2015). Emergent features and feature combination. *The Oxford Handbook of Perceptual Organization*, 88–107.
- Pomerantz, J. R., & Garner, W. R. (1973). Stimules configuration in selective attention tasks. *Perception & Psychophysics*, Vol. 14, pp. 565–569. doi:<https://doi.org/10.3758/bf03211198>
- Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic Gestalts. *Journal of Experimental Psychology: Human Perception and Performance*, 37(5), 1331–1349.
- Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, 3(3), 422–435.
- Pomerantz, J. R., & Schwaizberg, S. D. (1975). Grouping by proximity: Selective attention measures. *Perception & Psychophysics*, 18(5), 355–361.
- Spehar, B., & Halim, V. A. (2016). Created unequal: Temporal dynamics of modal and amodal boundary interpolation. *Vision Research*, 126, 97–108.
- Team, R., & Others. (2015). RStudio: integrated development for R. RStudio, Inc., Boston, MA URL <http://www.Rstudio.com>, 42, 14.
- Wagemans, J. (2018). Perceptual Organization. In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (pp. 1–70).
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217.
- Wallis, T. S. A., Funke, C. M., Ecker, A. S., Gatys, L. A., Wichmann, F. A., & Bethge, M. (2017). A parametric texture model based on deep convolutional features closely matches texture appearance for humans. *Journal of Vision*, 17(12), 5.
- Wei, N., Zhou, T., & Chen, L. (2018). Objective measurement of gestalts: Quantifying grouping effect by tilt aftereffect. *Behavior Research Methods*, 50(3), 963–971.
- Yssaad-Fesselier, R., & Knoblauch, K. (2006). Modeling psychometric functions in R. *Behavior Research Methods*, 38(1), 28–41.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.