Can Two Dots Form a Gestalt? Measuring Emergent Features with the Capacity Coefficient

Robert D. Hawkins

Joseph W. Houpt
Wright State University - Main Campus, joseph.houpt@wright.edu

Ami Eidels

James T. Townsend

Follow this and additional works at: https://corescholar.libraries.wright.edu/psychology

Part of the Psychology Commons

Repository Citation

This Article is brought to you for free and open access by the Psychology at CORE Scholar. It has been accepted for inclusion in Psychology Faculty Publications by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu, library-core scholar@wright.edu.
Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient

Robert X.D. Hawkins
Indiana University, Bloomington

Joseph W. Houpt
Wright State University

Ami Eidels
University of Newcastle, Australia

James T. Townsend
Indiana University, Bloomington

Author Note

This research took place while the first author was a student at Indiana University. He is now at Stanford University, Stanford, CA 94305, USA. This work was supported by National Institute of Mental Health Grant 05771710 awarded to J. T. T. We thank Devin Burns for helpful comments on the experiment reported, and Patricia Knapp for her assistance in data collection. Correspondence concerning this article should be addressed to Robert X.D. Hawkins, e-mail: rxdh@stanford.edu
Abstract

While there is widespread agreement among vision researchers on the importance of some local aspects of visual stimuli, such as hue and intensity, there is no general consensus on a full set of basic sources of information used in perceptual tasks. Gestalt theories place particular value on emergent features, which are based on the higher-order relationships among elements of a stimulus rather than local properties. Thus, arbitrating between different accounts of features is an important step in arbitrating between local and Gestalt theories of perception in general. In this paper, we present the capacity coefficient from Systems Factorial Technology (SFT) as a quantitative approach for formalizing and rigorously testing predictions made by local and Gestalt theories of features. As a simple, easily controlled domain for testing this approach, we focus on the local feature of location and the emergent features of Orientation and Proximity in a pair of dots. We introduce a redundant-target change detection task to compare our capacity measure on (1) trials where the configuration of the dots changed along with their location against (2) trials where the amount of local location change was exactly the same, but there was no change in the configuration. Our results, in conjunction with our modeling tools, favor the Gestalt account of emergent features. We conclude by suggesting several candidate information-processing models that incorporate emergent features, which follow from our approach.

Keywords: Perceptual Organization; Visual Perception; Workload Capacity
DOTS

Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient

Introduction

One of the central problems in vision science concerns the process by which raw visual input is organized into meaningful percepts that can ultimately be used to make decisions (Palmer, 1999; Kimchi, Behrmann, & Olson, 2003). Accounts of many perceptual tasks, such as visual search (Wolfe, 1994), object-recognition (Biederman, 1987), attention allocation (Moore & Egeth, 1998), categorization (Kruschke, 1992; Nosofsky, 1986) and memory (Luck & Vogel, 1997), rely on the notion of perceptual “features,” the elemental information that the perceptual system extracts from raw visual input and builds into percepts. Examples of proposed features range from basic physical properties like the hue, intensity, or location of an item in a scene to stimulus-specific properties like the eyes of a face or line orientations of block letters. Despite the importance of features in the psychological literature, there is no consensus about which of the infinite set of possible features are most informative, and how they interact in different contexts (Pinker, 1984; Schyns, Goldstone, & Thibaut, 1998; Treisman, 1988; Pomerantz & Portillo, 2012; Wolfe & Horowitz, 2004). This problem is also crucial for work in machine learning and computer vision, where systems must encode or learn a feature ‘vocabulary’ over which to make inferences (e.g. Austerweil & Griffiths, 2011; Blum & Langley, 1997).

To some extent, the debate over Gestalt processing is primarily a debate over features: when the perceptual system encounters a complex stimulus, does it break the stimulus into a set of local features that are subsequently pieced together into a percept, or does it act directly on higher-order (emergent or holistic) features that cannot be decomposed? We call the former view the local theory of features and the latter the Gestalt theory. In this paper, we present the capacity coefficient, $C(t)$, as a quantitative tool to arbitrate between these two views on features, and therefore as an approach to quantitatively test the predictions of Gestalt theory in general.

The capacity coefficient is a nonparametric measure of workload capacity from Systems Factorial Technology (SFT; Townsend & Nozawa, 1995). It measures change in performance as additional items are added to the display, allowing us to compare
processing of the ‘whole’ against the ‘sum of the parts’ in a principled way. The capacity coefficient is a potentially useful tool for the study of configural effects and Gestalt perception for a number of reasons: (i) it compares human performance to a well-established benchmark model (which we outline soon); (ii) it is based on the entire distribution of response times, not just means; and (iii) does not make assumptions about model-parameters, making it very general.

In previous studies, the capacity coefficient has been used to model configural effects in the word processing (Houpt, Townsend, & Donkin, 2014), face processing (Burns, Houpt, & Townsend, 2010), perceptual learning (Blaha, 2011), audio-visual integration (Altieri & Townsend, 2011), and visual search (Eidels, Townsend, & Pomerantz, 2008) domains. However, the complex, domain-specific nature of the stimuli used in these studies makes it difficult to generalize their conclusions to the overarching theory of Gestalt processing.

Consider, for example, the aforementioned study by Eidels et al. (2008). In their study, participants were presented with stimuli akin to those used by Pomerantz et al (1977): various combinations of a diagonal line (either left, \, or right, /) and a right angle (open either to the right, >, or to the left, <). Capacity was estimated from response-time data to inform about the underlying processing mechanisms. However, the complex interplay between basic features such as lines and angles and higher order features such as closure, symmetry, and even topological similarities between items in the set had made it hard to interpret each effect in isolation (in addition, these researchers were not interested in isolating effects of selected features).

In the current study we conducted a careful manipulation of the features posited by Gestalt theory by focusing on one of the simplest perceptual tasks in which the local and Gestalt views come into direct conflict: detecting a location change in a pair of dots. We developed a suitable redundant-target task to collect the reaction time data needed to compute capacity for different combinations of two of the lowest-level configural features posited by the Gestalt view in a pair of dots, Orientation and Proximity, and tested how they affect our model-informed capacity measure. Answering this question in an easy-to-control domain, where we can isolate features, may shed light on the processing mechanisms that underlie Gestalt perception in general.
Components or configurations?

Historically, there have been two main schools of thought on what constitutes a feature. The first supposes that a perceptual scene can be segmented into component pieces (e.g. the eyes, nose, and mouth of a face or the objects in a visual array), and the intrinsic physical properties of those pieces (e.g., location, color, brightness, size, spatial frequency) are the fundamental sources of perceptual information (e.g. Treisman & Gelade, 1980; Nosofsky, 1986; Luck & Vogel, 1997; Wolfe & Horowitz, 2004).

Typically, these features are characterized as static and able to be processed independently of one another, perceived as the same whether they appear together or in isolation (Garner, 1974; Rogosky & Goldstone, 2005). Local properties are easily extracted from a stimulus using image processing algorithms and are therefore implicitly utilized in template matching techniques, making local features popular and successful in computer vision (e.g. Brunelli & Poggio, 1993; Li & Allinson, 2008).

Another perspective comes from Gestalt studies demonstrating that people perceive a whole as different from the sum of its parts. For example, Tanaka and Farah (1993, 2003) showed that parts of a face are more easily recognized when presented in the context of a whole face than in isolation (but see Gold et al., 2012). Here, the most salient, fundamental sources of information (or features) are not local, but global (e.g. Navon, 1977; Pomerantz & Kubovy, 1986). They are present in the configuration or organization of the parts, and cannot be decomposed into a more fundamental set of independent features. They are therefore called emergent features, since adding new components can induce extra information beyond what is predicted by each component being processed in isolation, possibly through some unitization process (Hendrickson & Goldstone, 2009; Blaha, Busey, & Townsend, 2009).

The primordial examples of emergent features arose in the context of grouping. For example, when participants are presented with a lattice of dots where the horizontal distances between dots are smaller than vertical distances, they report that the induced horizontal lines are the most salient organization. When the horizontal distances are increased to a higher value than the vertical distances, however, the percept flips: participants report an organization into vertical lines. The properties of individual dots are
Figure 1. Example odd-quadrant stimuli adapted with permission from Pomerantz and Portillo (2011). (a) In the single-dot condition, participants were asked to select the quadrant that was different from the others. In this case, the correct response is the upper-left panel. (b) An uninformative context that is added to the single-dot stimuli to get (c), the composite stimuli. In general, responses on ‘single dot’ trials were found to be slower and less accurate than responses on composite Orientation trials, even though the additional dot added to create the Orientation feature provide no additional information on its own. Note that due to ‘false pop-out’ participants occasionally picked a quadrant different than the correct answer, because they felt it broke the symmetry (e.g., upper right quadrant in panel a).

subsumed by their overall organization, and the phenomenology is controlled by a small set of parameters (Kubovy & Gepshtein, 2003).

Pomerantz and the Odd-Quadrant Task

Many further examples of emergent features have been discovered outside the grouping domain as well. Early evidence for the salience of emergent features in perception came from an odd-quadrant paradigm (see Figure 1). In its original formulation, participants were presented with a four-panel display with three of the panels containing the same stimulus and the fourth containing a different stimulus (Pomerantz et al., 1977).
The participant was asked to pick the ‘odd-quadrant’ as quickly and accurately as possible. In some trials, the ‘component’ appeared in isolation. For instance, a single dot was presented at the bottom left of three panels and at the top or mid-left of the fourth panel (see Figure 1a). In other trials, some non-informative context (Figure 1b) was added to all quadrants to form a composite stimulus (Figure 1c).

This context was non-informative in the sense that no local information about it could be used to distinguish the odd-quadrant. However, it often impacted reaction times and accuracy in the composite condition. When the configuration induced by the context improved performance, it was called a *configural-superiority* effect; when it negatively affected performance, it was called a *configural-inferiority* effect. Over the years, Pomerantz and colleagues (Pomerantz, 1983; Treisman & Paterson, 1984; Pomerantz & Portillo, 2011) have postulated a number of emergent features for lines and dots which could account for these results.

In the present work, we extend the modeling approach for identifying configurality introduced by Eidels, Townsend, and Pomerantz (2008) by addressing the simplest possible case in which emergent features can become salient in visual perception: a pair of dots. An isolated dot is defined solely by its spatial coordinates in the plane. When additional dots are added, their $x$ coordinate and $y$ coordinate provide additional sources of information, but new features also emerge from the relationship *between* the dots. Following Pomerantz and Portillo (2011), these new features include Proximity (distance between dots), Orientation (angle of implicit line between dots), Linearity (whether three dots or more appear along the same imaginary line), and Surroundedness (if one dot is in the interior of an imaginary polygon formed by at least three other dots).

Their study compared response times across various conditions, yet important information can be lost by not considering the entire RT distributions (Townsend, 1990). We apply the sophisticated RT machinery of Townsend and colleagues, taking advantage of the full RT distributions to calculate the capacity coefficient in order to learn about the underlying properties of the cognitive system when it extracts emergent features. We focus in particular on the emergent properties of Orientation and Proximity. To apply the capacity coefficient to this problem, we developed a suitable redundant-target task to
Figure 2. Stimuli and procedure in change detection task. (a) the three classes of reference stimuli, containing one or both of the channels of local information. (b) the sequence of displays in a ‘control’ trial. Because the dots changed location in the second frame, the participant should respond ‘change’. (c) the sequence of displays in a ‘configural’ trial. Both channels provide the same amount of location change information, but there is also a change in the orientation of the dots, which Gestalt theories predict will lead to more efficient processing.

collect the reaction time distributions and accuracies associated with different combinations of features.

**Redundant-Target Task**

Participants were presented with a reference display showing either a stimulus to the left of the center (L only), a stimulus to the right of center (R only), or stimuli in both positions (R & L; see Figure 2a). The reference screen was followed by a brief masking
stimulus, then the participant was shown a display in which the dot(s) were in either the same location as the reference or a different location (Figure 2b,2c). The masking length was calibrated to the shortest level at which pilot participants no longer reported apparent motion cues.

Participants were asked to respond whether or not the dot(s) were in the same location before and after the mask. When two dots were displayed in the reference screen, either both dots moved or neither moved. Trials in which both dots were in a different position than the reference contain redundant information; noticing any one of the components moving by itself is sufficient to complete the task, but if the Gestalt account of emergent features is correct, then we predict that when both dots are present, additional configural information is available to participants. Thus, for the study of holistic or Gestalt effects, it is instructive to compare performance when components appear together (R & L) against baseline performance expected when they appear in isolation (L only or R only).

There are three main advantages that a redundant-target task holds over the odd-quadrant task introduced by Pomerantz, Sager, & Stoever (1977), which requires a full visual search with multiple eye fixations. First, the requirement of multiple fixations can confound the measurement of reaction time. Second, the odd-quadrant task is known to induce a ‘false pop-out’ effect for certain stimuli (Orsten & Pomerantz, 2012), in which another level of configural grouping is made across separate quadrants. This effect interferes with the lower-level grouping phenomena under investigation. For instance, in Figure 1, a configural-inferiority effect was found, despite the change in Orientation, because participants chose the quadrant that was not ‘pointing toward the center’ and therefore breaking the higher-order symmetry. Our task obviates the first problem by including the entire display within the parafoveal region, such that only a single fixation is required, and avoids false pop-out effects by limiting the presentation to a single component or configuration on the screen at a time. Third, the design lends itself to analyses of data using Systems Factorial Technology, which we present next.
Systems Factorial Technology

Predictions for reaction time distributions under local and Gestalt theories can be generated using the capacity coefficient, a key tool from Systems Factorial Technology (SFT; Townsend & Nozawa, 1995; Wenger & Townsend, 2001; Townsend & Wenger, 2004; Houpt & Townsend, 2012), a mathematical framework for revealing the underlying properties of cognitive processes. It provides rigorous, non-parametric tests for architecture, stopping rule, inter-channel dependencies, and workload capacity using a synthesis of response time methodologies from factorial approaches and redundant-target designs.

Conceptually, the capacity coefficient measures the efficiency of a cognitive process relative to the baseline prediction of a parallel race model, which formalizes the situation in which local information from each source (i.e. each feature) is processed independently. In the current design, a change in the position of the left dot (\(L\)) is one source of local information and a change in the right dot (\(R\)) is another source of local information. Then the capacity coefficient is defined as the ratio:

\[
C(t) = \frac{H_{LR}(t)}{H_L(t) + H_R(t)}
\]

(1)

where \(H_{LR}\) is the cumulative hazard function derived from the response time distribution when both sources of information indicate a target simultaneously and \(H_L, H_R\) are the cumulative hazard functions derived from the response time distribution when each source is presented in isolation and indicates a target. The cumulative hazard function represents the likelihood of the response process terminating at time \(t\) given that it has persisted until time \(t\). It can be derived as the negative log of the survivor function \(S(t)\), which is simply \(1 - F(t)\), where \(F(t)\) is the empirical response time CDF. The capacity coefficient is typically used as an absolute measure categorizing a process as limited, unlimited, or supercapacity depending on whether \(C(t)\) is less than, equal to, or greater than 1. Here we use it instead as a sensitive relative measure across conditions. Following Houpt and Townsend (2012) we derive a capacity z score, which is a convenient summary statistic for \(C(t)\). Because \(C(t)\) and the capacity z score are different transformations of the same data, we use the terms interchangeably in the text.
Overview of the Experiments

We present three experiments in which the capacity coefficient is used to conduct a critical test of local and Gestalt theories. Experiments 1 and 2 test the local features of dot location against the emergent feature of Orientation. While they use the same stimuli, they differ in the block structure used to present these stimuli. This allows us to test the robustness of our measure with respect to details of the experimental procedure, and to replicate our overall results. Experiment 3 proceeds to test the local features of dot location against the emergent feature of Proximity.

All three experiments used a $2 \times 2$ within-subject factorial design manipulating (1) the presence or absence of configural cues in redundant-target trials and (2) the presence or absence of an explicit line connecting the dots. Note that unlike previous SFT studies, which employ a double factorial paradigm, we do not manipulate the salience of configural cues, just their presence or absence. This reserves the second dimension of the factorial design to test the presence of a line. In the redundant-target trials, the components either moved in the same direction to preserve Orientation (“control”; e.g., both dots moving up, as in Figure 2b) or moved in opposite direction to induce a change in emergent feature (“configural”; 'Figure 2c). In both cases, there is the same amount of local information available, since the components move the same amount in either direction. Hence, the local theory predicts that the capacity coefficient will be the same in control and configural trials. The Gestalt theory, on the other hand, predicts that the capacity coefficient will be larger in the configural trial, since the change in emergent feature serves as an additional source of information.

Since the orientation and length of an explicit line is canonically considered a local feature, the second manipulation compares the information provided by the implicit (or imaginary) line between the dots to the information provided by an explicit line. The local theory predicts an interaction: capacity should be higher in the ‘explicit line’ condition than the ‘implicit line’ condition when configural cues are available, since additional information about orientation and length is available. The Gestalt theory predicts no effect of the line, since the physical features provided by the line were already present as emergent features in the dots. Thus, the application of SFT and specifically the capacity coefficient
provides a critical test for the role of emergent features and therefore of Gestalt perception.

**Experiment 1**

**Methods**

**Participants.** Twenty-one paid individuals between the ages of 18 and 24 were recruited from the Indiana University student population to participate in two 50 minute sessions. Six participants were removed from the study after their first session due to high error rates. Of the participants that completed both sessions, ten were female, five were male, and all had normal or corrected-to-normal vision. In accordance with the Declaration of Helsinki, the procedures were approved by local IRBs and signed consent forms were obtained from individual participants before the experiment.

**Materials.** All stimuli were created using the Script-Fu plugin for Gimp and presented using DMDX (Forster & Forster, 2003) on a 17" inch Viewsonic CRT monitor at 1024 x 768 resolution with a 75GHz refresh rate and luminance of 150 cd/m². The dots in the stimuli were grey with 50% the luminance of the background (hex: 7F7F7F) and with a diameter of 0.34° in visual angle, at a sitting distance of approximately 70 cm. Responses were collected using a button box connected with a Measurement Computing PCI-DIO24 Interface Card.

We used four different classes of stimuli, in which the distance between the dots’ inner contours was always held at a constant visual angle of 1.10° to avoid possible confounds with Proximity. Figure 3 displays the possible positions of each dot. Note that each possible target position (denoted by the filled circles) is an equal distance away from the reference position (open circles). The green circles correspond to possible positions for the left channel, and blue circles correspond to possible positions for the right channel. The green and blue colors are only used for illustration purposes in the figure. For each of the following classes of two-dot stimuli, corresponding single-dot stimuli were presented to collect response times for the isolated components:

1. **Configural, no line:** Each dot is 0.74° of visual angle away from its initial positions to a point opposite the other on a circle (Figure 3). The implicit line between them is
Figure 3. (a) Possible locations of dots in Experiments 1 and 2. Note that all possible locations for each dot are the same distance away from the reference location, forming an equivalence class under the metric of Euclidean distance. Single-dot stimuli were presented for every position. (b)-(c) Configural stimuli are formed by moving the dots to antipodal points on the circle (i.e. Green 2, Blue 3 or Green 4, Blue 1), holding Proximity constant. (d)-(g) For each point on the circle, a control stimulus can be formed by adding a new position on the same horizontal line.

approximately 60° away from the horizontal. There are two variations of this stimulus – one where the left dot goes up and the right dot goes down (Green 2, Blue 3; panel b) and another where the left dot goes down and the right dot goes up (Green 4, Blue 1; panel c). The appropriate degree of configural change was chosen
using the results of a pilot study measuring the $d'$ for different levels of Orientation (Supplemental Figure S1).

2. **Control, no line**: Both dots are still the same distance from the reference point as in the configural conditions, but move in the same direction (Green 1, Blue 1; Green 2, Blue 2, etc.) Thus, the implicit line between them remains horizontal and there is no change in configural features.

3. **Configural, line present**: Like the other configural condition, but on double-dot trials, a line connected the two dots.

4. **Control, line present**: Like the other control condition, but a line connected the two dots.

**Procedure.** The sequence of displays in a trial is shown in Figure 2b,c. On each trial, a fixation cross appeared in the center of the screen for 200 ms, followed by a blank display for 27ms. On single-dot trials, a blue square was presented 0.72° of visual angle to either the left or the right of the center fixation. On double-dot trials, blue squares were presented in both positions simultaneously. This reference screen remained for 120ms and was then masked for 240ms by one of five randomly generated Gaussian noise patterns. The probe stimulus was displayed for 120ms, followed by a blank screen for 1880ms. Response times were calculated from stimulus onset.

At the beginning of the session, participants were instructed to press one button (‘no change’) if the probe dots were in the same locations as the reference squares and another button (‘change’) if the probe dots were in a different location. Participants received feedback on negative responses and time-outs for 20 practice trials at the beginning of each session, but did not receive any feedback for the remainder of the session.

Each subject participated in two 50-minute sessions of 960 trials per session. One session contained exclusively ‘line’ trials, while the other contained exclusively ‘no line’ trials. Configural and control stimuli were split into separate blocks. Within each session, however, there were three contiguous blocks of ‘configural’ trials and three contiguous blocks of ‘control’ trials, with optional rest breaks between blocks. The corresponding
Figure 4. (a) Mean response times and (b) accuracy for each condition in Experiment 1 (Orientation with separate blocks for configural and control trials). Configural trials differed from the reference in Orientation as well as the location of each element. In control trials, both elements were in a different location than the reference squares, but the Orientation was the same. In distractor trials, both elements were in the same location as the reference squares. Error bars indicate 95% highest density intervals of the posterior. (c) Mean capacity z-scores for each condition in Experiment 1. Positive numbers indicate better than the unlimited capacity, independent parallel baseline, while negative numbers indicate worse than the baseline. In general, higher numbers indicate more efficient responding. Error bars indicate 95% highest density intervals of the posterior.

‘single-dot’ trials were mixed into each block. The ordering of sessions and the ordering of ‘configural’ and ‘control’ block sets within each session was counterbalanced across participants. The distribution of stimuli was again chosen to balance the conditional probabilities: there was a 25% chance of no change (negative response), 25% chance of a double-dot change (positive response) and 50% single-dot trials evenly spread over all possible locations. The two variations of configural trials and four variations of control trials were evenly distributed.
Results

Bayesian ANOVAs (Rouder, Morey, Speckman, & Province, 2012) were used to analyze mean correct response times and accuracy. Within this framework, we calculated Bayes Factor (BF) for each effect of interest, with the convention that BF>10 is strong evidence and BF>100 is decisive evidence (see Jeffery, 1961). BF<3 is weak evidence, and BF<1 is ‘negative’ evidence, in favor of the null model. Figure 4 shows the mean response times (a) and accuracies (b) for trials in which two dots were present along with the 95% highest density intervals of the posterior. The highest density intervals (HDI) is the smallest interval of the posterior containing 95% of the density.

The analysis of correct response times for two dot stimuli indicated main effects of configuration (BF = 2.3 · 10^70) and of lines (BF = 1.5 · 10^22) and was nearly equivocal with respect the presence of an interaction (BF = .53). In the accuracy data, there was very strong evidence against an interaction between the configuration and the presence of lines (BF = .025). There was decisive evidence for main effects of configuration (BF = 9.8 · 10^19) and lines (BF = 1.2 · 10^5).

For capacity we use Houpt and Townsend (2012) z score as a summary statistic for C(t) that can be subjected to inferential tests. Capacity z scores of zero indicate unlimited capacity. Capacity z scores could also be positive or negative, indicating super- or limited-capacity, respectively. The Bayesian ANOVA on capacity Z scores (shown in Figure 4(c)) indicated that the most likely model includes a main effect for only configuration (BF = 1.2 · 10^6 over a subject only model). Evidence against including an additional main effect of the line was again weak (BF = 0.34) and there was substantial evidence of the configural main effect only model relative to the model with both main effects and an interaction (BF = 5.4). The mean posterior advantage of configural over control on the capacity z-scores was 3.15 (HDI = [2.14, 4.12]). The mean posterior difference between capacity z-scores without lines and with lines was −0.43 (HDI = [−1.29, 0.47]).

Participants were generally quite limited capacity, with a group average capacity z-score of −3.57 (HDI = [−4.43, −2.65]. Nonetheless, there remained at least a few participants who had capacity z-scores that indicated super-capacity in a configural condition (see Table 1).


Discussion

The two channels contributed the same amount of location information in each condition, but the configuration of the dots drastically affected mean response time, accuracy, and the capacity coefficient $C(t)$. When a source of configural information was present, participants performed much more efficiently on the whole, compared to the sum of its parts, as measured by $C(t)$. This effect was predicted by the Gestalt view of features, but not the local view of features.

Including an explicit line between the dots, which canonically has the physical feature of Orientation, also impacted response times and accuracy, but in the opposite direction; response times were higher when lines were present and accuracy was lower. The data were not as clear with respect to an effect of the lines on the capacity values. One explanation of this result would be that because the location of the dots already contains all of the Orientation information, the addition of the line offers no additional advantage, but instead limits performance by using up additional processing resources.

It is possible, however, that these results can be accounted for by the block structure of configural and control trials. By isolating stimuli from each condition in separate blocks, participants could have been biased to focus on the information provided by obvious Orientation differences to the exclusion of the location information in Orientation trials. To address this concern and replicate our results, we ran a second experiment where the blocks were mixed together, which does not allow participants to use different processing strategies a priori.

Experiment 2

Methods

Participants. Twenty paid individuals between the ages of 18 and 26 were recruited from the Indiana University community to participate in two 60 minute sessions. Five participants were removed from the study after their first session due to unacceptably high error rates of 30% or greater. Of the participants that completed both sessions, fourteen were female, one was male, and all had normal or corrected-to-normal vision. In
accordance with the Declaration of Helsinki, the procedures were approved by local IRBs and signed consent forms were obtained from individual participants before the experiment.

**Materials.** All equipment and stimuli were the same as in the previous experiment.

**Procedure.** The procedure was identical to Experiment 1 except that configural and control trials, along with their corresponding single-dot trials, were mixed together and presented in random order across 4 blocks with short rest breaks between blocks. Also, instead of 960 trials per 50-minute session, we used 1152 trials per 60-minute session. Again, one session contained only ‘line’ trials and the other contained only ‘no line’ trials. The distribution of stimuli was chosen to balance the conditional probabilities of presentation: there was a 25% chance of no change, the positive responses distributed with 12.5% double-dot ‘configural’, 12.5% double-dot ‘control’, and the other 50% were single-dot trials evenly spread over all possible locations.

**Results**

Figure 5 shows the mean response times (a) and accuracies (b) for trials in which two dots were present along with the 95% highest density intervals of the posterior. The analysis of correct response times for two dot stimuli indicated main effects of configuration (BF = 2.7 · 10^{69}) and of lines (BF = 2.3 · 10^{3}) and was nearly equivocal with respect the presence of an interaction (BF = .51). In the accuracy data, there was decisive evidence for an interaction between the configuration and the presence of lines (BF = 104). When the interaction was disregarded, there was decisive evidence of a main effect of configuration (BF = 2.0 · 10^{65}) and nearly equivocal evidence against a main effect of lines (BF = .53).

Capacity Z scores were again calculated following Houpt and Townsend (2012) for each participant in each condition and are shown in Figure 5(c). Those values were then compared using a Bayesian ANOVA across the configural-control manipulation and the implicit-explicit line manipulation. The most likely model included only a main effect of configuration (BF = 6.8 · 10^{12} over a subject only model) however there was only weak evidence for leaving out an additional main effect of the line (BF = 2.8). The analysis did indicate substantial evidence for the configuration only model when compared to a model including both lines and an interaction (BF = 8.0). The mean posterior advantage of
Figure 5. (a) Mean response times and (b) accuracy for each condition in Experiment 2 (Orientation with mixed configural and control blocks). Configural trials differed from the reference in Orientation as well as the location of each element. In control trials, both elements were in a different location than the reference squares, but the orientation was the same. In distractor trials, both elements were in the same location as the reference squares. Error bars indicate 95% highest density intervals of the posterior. (c) Mean capacity z-scores for each condition in Experiment 2. Positive numbers indicate better than the unlimited capacity, independent parallel baseline, while negative numbers indicate worse than the baseline. In general, higher numbers indicate more efficient responding. Error bars indicate 95% highest density intervals of the posterior.

Configural over control on the capacity z-scores was 5.83 (HDI= [4.77, 6.91]). The mean posterior difference between capacity z-scores without lines and with lines was −0.387 (HDI= [−1.32, 0.567]).

The grand mean for the capacity z scores at the group level was negative, −4.94 (HDI= [−5.96, −3.94]), implying limited capacity. However, in the configural condition, there was some variability across participants, with several participants’ data indicating super capacity (positive z score) or indistinguishable from unlimited capacity (z ≈ 0; see Table 1).
Discussion

We replicated the results of Experiment 1 with configural and non-configural trials intermixed. This ruled out the possibility that participants only performed at higher capacity in the presence of an Orientation cue because they were primed to expect it by the block composition. The likelihood that the upcoming target would be identifiable only using differences in location was equivalent to the likelihood that it could be identifiable using differences in configuration, so participants could not have successfully adopted a strategy of ignoring location information.

Although the overall pattern of results matches Experiment 1 almost perfectly, there were some differences. First, the magnitude of the capacity advantage for configural trials over control trials was larger in Experiment 2 (5.83 compared with 3.15). This is likely due to the relatively worse capacity for the control trials in Experiment 2 because the mean capacity $z$ scores for the configural trials are nearly identical across the two experiments. This drop in efficiency on control trials may be due to participants giving processing priority to detecting a configural cue in the mixed condition, then checking location if the configural cue is absent. In Experiment 1, when the control trials were in their own block, participants would not gain any advantage from checking for configural differences because there were not any.

A second difference between Experiments 1 and 2 was that there was clear evidence for an interaction between the lines and the configuration in Experiment 2, although only in the accuracy. It is clear from Figure 5(b) that the interaction has a fairly small magnitude, so we will not dwell on it here beyond noting that it seems to be driven by an increase in accuracy for the target present trials due to the additional line context and a decrease in the distractor trials with the addition of lines.

Since the choice of ‘mixed’ or ‘separated’ block designs did not affect our conclusions, we proceeded to test the emergent feature of Proximity using the ‘separated blocks’ design.
Experiment 3

Methods

Participants. Twenty-four paid individuals between the ages of 20 and 32 were recruited from the Indiana University community to participate in two 50 minute sessions. Two participants dropped out of the study after their first session, and six were removed from the study after the first session due to unacceptably high error rates of 30% or greater. Of the sixteen participants that completed both sessions, thirteen were female, three were male, and all had normal or corrected-to-normal vision. In accordance with the Declaration of Helsinki, the procedures were approved by local IRBs and signed consent forms were obtained from individual participants before the experiment.

Materials. Using the same settings as Experiments 1 and 2, we created two new classes of stimuli, with the dots always lying on a horizontal axis (0°) to avoid confounds with the emergent feature of Orientation. Figure 6a displays the possible positions of each dot. Note again that each possible target position (denoted by the filled circles) is an equal distance away from the reference position (open circles). For each of the following classes of two-dot stimuli, corresponding single-dot stimuli were presented to collect response times for the isolated components:

1. Configural, no line: Each dot is displaced by 0.17° of visual angle away from its initial position toward the edge of the display (Green 1, Blue 2; Figure 6b). This expands the initial distance between reference points by a factor of 1.72, thereby inducing a change in the emergent feature of Proximity. The appropriate degree of configural change was chosen using the results of a pilot study measuring the $d'$ for different levels of Proximity change (Figure S2).

2. Control, no line: The individual dots are displaced the same amount as in the configural condition, but in the same direction (Green 2, Blue 2 and Green 1, Blue 1; panels c and d, respectively). Thus, the Proximity between the dots remains constant while the individual ‘channels’ contain the same information about location change.

3. Configural, line present: Like the other configural condition, but on double-dot
Figure 6. (a) Possible locations of dots in Experiment 3. Green dots denote possible locations for the left dot, and blue dots denote possible locations for the right dot. Note that all possible locations for each dot are the same distance away from the reference location, forming an equivalence class under the metric of Euclidean distance. Single-dot stimuli were presented for every position. (b) Configural stimuli are formed by moving the dots in opposite directions (Green 1, Blue 2), increasing the distance between them by a factor of 1.72. (c)-(d) For both of these outer positions, a control stimulus was formed by moving the opposite dot such that the distance between the reference dots was preserved.

4. **Control, line present**: Like the other control condition, but a line connected the two dots.

**Procedure.** The task and protocol were identical to Experiment 1.

**Results.** Figure 7 shows the mean response times (a) and accuracies (b) for trials in which two dots were present along with the 95% highest density intervals of the posterior. The analysis of correct response times for two dot stimuli indicated main effect of configuration (BF= $4.6 \cdot 10^7$) but very strong evidence against main effect of lines (BF= 0.026) and substantial evidence against a full model including an interaction relative the model only including a main effect of configuration (BF=.11). In the accuracy data, there was decisive evidence for an interaction between the configuration and the presence
of lines relative to the main effects only model ($\text{BF} = 1.4 \cdot 10^{52}$). When the interaction was disregarded there remained decisive evidence of main effects of configuration ($\text{BF} = 4.0 \cdot 10^{43}$) and lines ($\text{BF} = 5.4 \cdot 10^{4}$).

While overall error rates were lower than 30% for all sixteen participants who completed the study, three participants had error rates equal to or worse than chance when restricted to trials from one or more of the four conditions (e.g., the configural trials with lines). Since the capacity coefficient analysis only uses response times from correct responses, this potential difference in response thresholds could bias comparisons between conditions. For the following analysis, we only report the thirteen participants with above chance accuracies in all conditions.

The Bayesian ANOVA on capacity $Z$ scores (shown in Figure 7(c)) indicated the most likely model included both main effects and an interaction ($\text{BF} = 1.3 \cdot 10^{8}$ over the subject only model). There was substantial evidence for the full model over the next best model, which included only main effect of configuration ($\text{BF} = 9.9$) and strong evidence over the third best model, which included both main effects ($\text{BF} = 12$).

The mean marginal posterior advantage of configural over control on the capacity $z$-scores was $4.44$ (HDI $=[3.47, 5.41]$). The mean posterior difference between capacity $z$-scores without lines and with lines was $-0.778$ (HDI $=[-1.73, 0.176]$).

Participants were again generally limited capacity, with a group average capacity $z$-score of $-1.56$ (HDI $=[-2.92, -0.0821]$). In one condition, configural without lines, performance was super-capacity at the group level ($Cz = 1.95$, HDI $= 0.0193, 3.45$). There were also a few participants whose individual data indicated super capacity in the configural condition with lines, but none in either of the control conditions (see Table 3).

**Discussion.** First, the capacity coefficient measure is again larger in the configural condition than the control condition, indicating that Proximity is indeed an emergent feature providing additional information above and beyond the contribution of the individual dot locations. The accuracy interaction was again present and still had a limited effect size, however the crossover from Experiment 2 is not evident in these data. Instead,

---

1We have also run the analyses including the three low accuracy subjects. The magnitudes of the reported values were slightly different but none of the conclusions changed.
Figure 7. (a) Mean response times and (b) accuracy for each condition in Experiment 3 (using Proximity and separate configural and control blocks). Configural trials differed from the reference in Proximity as well as the location of each element. In control trials, both elements were in a different location than the reference squares, but the Proximity was the same. In distractor trials, both elements were in the same location as the reference squares. (c) Mean capacity $z$-scores for each condition. Positive numbers indicate better than the unlimited capacity, independent, parallel baseline, while negative numbers indicate worse than the baseline. In general, higher numbers indicate more efficient responding. In all panels, error bars indicate 95% highest density intervals of the posterior.

the effect seems to be driven by a larger magnitude drop in the distractor correct rejections between the ‘line’ and ‘no lines’ conditions.

Unlike the previous two experiments focusing on Orientation, however, we also see an interaction between the line manipulation and the configural condition on the capacity $z$-scores. In Experiments 1 and 2 there was weak evidence against an effect of lines and substantial evidence against an interaction. The benefit of the configural cue of Proximity compared to the control condition, measured in terms of capacity, was greater when the two dots were not connected by a line. The presence of a line appears to inhibit the contribution of configural information. This is the opposite of the interaction predicted by the local theory, and also by the literature on redundant signals, which suggest that the presence of additional explicit cues should improve detection.
The most likely account of this interaction is through the Gestalt phenomenon of ‘element connectedness’ (Palmer & Rock, 1994), where connecting two dots by a line segment strengthens their tendency to be grouped together. If the dots are grouped together more strongly, the true increase in Proximity is counteracted, yielding a weaker effect. Interestingly, element connectedness does not seem to affect performance in the control condition, where Proximity stays constant. While there have been rigorous psychophysical studies of the strength of grouping by Proximity as a function of distance (Kubovy, Holcombe, & Wagemans, 1998), there is no psychophysical data about the impact of element connectedness on the perception of Proximity. Since element connectedness has only been discussed in the context of grouping, we have no a priori expectations about its strength when only two elements are present. This is an example where multiple Gestalt principles come into conflict, which remains an important direction for further investigation.

**General Discussion**

In all three experiments, we used the capacity coefficient as a diagnostic measure to show that the Gestalt theory of features provides a better explanation of the data than the local theory. When there is a change in emergent features of Orientation or Proximity, the perceptual system experiences gains in efficiency that cannot be accounted for in terms of how it processes the parts. Moreover, the presence of an explicit line does not provide any information not already present in emergent features between dots, and in the case of Proximity actually inhibits processing. This comparison of the whole against the sum of the parts has been at the core of Gestalt theory since its inception, and the capacity coefficient makes this comparison rigorous, in terms of response-time distributions and a well specified benchmark model.

We now turn to some details of our results that raise interesting questions for future work. First, note that while \( C(t) \) was much larger on configural trials than on control trials, there was still high variation across individuals. This is troubling for a natural characterization of configurality as high absolute performance relative to the parallel independent race model. Often, participants were still performing with *limited capacity*
DOTS

\((C(t) < 1)\), in the configural condition, which implies less efficiency than if local information was processed independently. One explanation for this effect is the existence of attentional factors that may interfere with processing and generally reduce workload capacity. However, because any such factors affect all trials evenly, it does not affect our comparison with control trials. Hence, when modeling the contribution of emergent features, we should be careful to measure *degrees* of configurality – as we did here – instead of making an absolute comparison.

If the model containing only local information does not account for the data, we are left with the question of what model *is* appropriate? Here, we propose several possibilities. Further work is needed to distinguish among them.

1. **Additional Channels**: Emergent features like Orientation and Proximity could constitute separate sources of information and “race” in parallel against local information coming from the individual dots. Under this theory, configural effects appear when channels containing information higher-order features overpowers the channels containing local information in that race. It has recently been suggested that topological similarity may play such a role (Pomerantz, 2003; Eidels et al., 2008). This model also has the advantage of generalizing easily to more complex stimuli (e.g. three or more dots), with additional higher-order features like co-linearity or symmetry successively overpowering lower-order features. Its potential scalability makes it a promising contender for implementation in a computer vision system. However, other properties of the race remain unclear, such as the degree of facilitatory and inhibitory interaction between channels (Eidels, Houpt, Altieri, Pei, & Townsend, 2011).

2. **Configuration-First Processing**: The visual system first takes holistic features like Orientation or Proximity into account and only examines local information if the holistic features are not informative enough to make the decision. There was some support for this model in the mixed design of Experiment 2. Recall that we found a decrease in processing efficiency for control trials when mixed together with configural trials, as compared to the same trials in Experiment 1, where participants
could plausibly use a “location-only” strategy. The “configuration-first” model could be more carefully tested against the “additional channels” model by designing new stimuli in which Orientation or Proximity changes the same amount as in the present study, but the degree of location change of the individual dots is much larger. Top-down processing predicts that there would be no difference in the results, since the information from individual dots would not be considered. However, the additional-channels model predicts that given enough of a boost, the channel containing local information could overpower the configural channel.

3. Coactivation: The location information from each dot could pool into a common channel that takes featural information into account (Miller, 1982; Colonius & Townsend, 1997). This model is theoretically appealing since it specifies an internal transformation by which local, physical information is transformed into higher-order percepts. However, our findings that stimuli containing emergent features are processed with limited capacity rule out this model, which predicts super capacity (Townsend & Nozawa, 1995). Coactivation was also recently ruled out as a viable model for configural processing because of its inability to predict behavior in trials containing distractors (Eidels et al., 2008).

In conclusion, we have presented strong evidence from a new experimental task, with inferences drawn using the powerful modeling approach of the capacity coefficient, that the simple emergent features of Orientation and Proximity between two dots confers a benefit to efficiency above and beyond the contribution of its component parts. Although these features are not local, physical properties of the stimulus, their contribution is indistinguishable from (and sometimes more efficient than) the local information provided by the orientation and length of an explicit line. By illustrating the critical role that the capacity coefficient played in our formalization and testing of Gestalt and local theories in this simple domain, we set the foundation for further work systematically investigating the processing of emergent features.
References


Table 1
Results from Experiment 1 broken down by participant and condition. Z gives the Z-score for the capacity coefficient statistic, with negative values implying limited capacity (comparable to \( C(t) < 1 \)) and positive values implying super capacity (comparable to \( C(t) > 1 \)). Note that several participants performed at unlimited or super capacity levels on configural trials, but all participants were significantly limited capacity on control trials.
Table 2

Results from Experiment 2 broken down by participant and condition in the same format as Experiment 1.
Table 3

Results from Experiment 3 broken down by participant and condition in the same format as Table 1.
Figure S1. Mean RT and $d'$ plotted as a function of the objective intensity of change between (dot) Orientation values in our change-detection task. All seven participants were presented five blocks of 200 trials each, where each block was composed of change-detection tasks at a fixed level of intensity. Block order was randomized, ‘change’ trials were evenly interspersed with ‘no change’ trials, and ‘change’ trials were evenly divided into positive and negative Orientation changes. We selected the intensity level of 60° for use in Experiments 1 and 2, both to maximize $d'$ and to set the expected RT at a comparable level to that of the redundant-target Proximity stimulus, shown in Figure S2.
Figure S2. Mean RT and d’ plotted as a function of the objective intensity of change between Proximity values in our change-detection task. The procedure was identical to the Orientation piloting in Figure S1, with eight participants. Note that a logarithmic scale was used for the intensity values to cover the space more efficiently. Intensity values on the x-axis indicate the coefficient by which the initial distance between dots was multiplied to get the ’change’ target. We selected the scaling coefficient of $\times 1.72$ for use in Experiment 3, in order to most closely match the mean response time of the chosen value of Orientation above and to set d’ so that expected error rates in Experiment 3 would be as low as possible.