Vision Research xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

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

Robert X.D. Hawkins [a,*], Joseph W. Houpt [b], Ami Eidels [c], James T. Townsend [a]

[a] Indiana University, Bloomington, United States
[b] Wright State University, United States
[c] University of Newcastle, Australia

A R T I C L E I N F O

Article history:
Received 27 November 2014
Received in revised form 25 April 2015
Available online xxxx

Keywords:
Perceptual organization
Visual perception
Workload capacity

A B S T R A C T

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 or how they are processed. 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.

1. 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 (Kimchi, Behrmann, & Olson, 2003; Palmer, 1999). Accounts of many perceptual tasks, such as visual search (Wolfe, 1994), object-recognition (Biederman, 1987), attention allocation (Moore & Egeth, 1998), categorization (Kruschke, 1992, 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; Pomerantz & Portillo, 2012; Schyns, Goldstone, & Thibaut, 1998; Treisman, 1988; 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 that derives from an extensive body of work using

☆ *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.

* Corresponding author.
E-mail address: rxdh@stanford.edu (R.X.D. Hawkins).

http://dx.doi.org/10.1016/j.visres.2015.04.019
0042-6989/© 2015 Published by Elsevier Ltd.

stochastic processes to model reaction time distributions under different information-processing constraints. This measure is part of a set of related tools for assessing the architecture, stopping rule, and independence of channels, known collectively as Systems Factorial Technology (SFT; Townsend & Nozawa, 1995). The capacity coefficient measures change in performance as additional items are added to the display, giving a principled way of integrating reaction time distributions about the 'parts' to make predictions about the 'whole'. Thus, the capacity coefficient can be directly interpreted as a measure of *processing efficiency*, which can be compared to the performance of certain well-defined benchmark models such as the parallel race model (Miller, 1982, 1991).

In brief, we define the capacity coefficient in terms of processing times for two sources of information: A and B presented either together or in isolation. Using the response times produced when the sources are presented in isolation, we estimate the predicted response time distribution when presented *together* assuming a parallel race model (i.e., A and B are processed in parallel at the same rate they would be if they were in isolation and a response occurs as soon as either of A or B are finished processing). In the capacity coefficient, we carry out the comparison between predicted performance and observed performance (with both sources present) in terms of the cumulative hazard function, $H(t) = -\log(F(t))$, where $F(t)$ is the cumulative distribution function. In these terms, the ratio of the redundant-target hazard function (the 'whole') and the sum of the individual channel hazard functions (the 'parts') should be equal to one. Ratio values below one indicate worse performance than a race model while above one indicates better performance than a race model. Further details of the measure are given below in the *Systems Factorial Technology* section.[1]

$$C(t) = \frac{H_{AB}(t)}{H_A(t) + H_B(t)} \tag{1}$$

The application of a model-based approach in general, and an approach based on the capacity coefficient in particular, yields a number of advantages for the quantitative study of emergent features and Gestalt perception:

(i) Framing the problem of configural perception in terms of workload capacity supplements and enriches the vocabulary typically used to characterize Gesalt phenomena. This is in line with the larger push toward theory-driven methodology in the psychological sciences: by considering the capacity coefficient as a theoretical construct, we can design a targeted, well-controlled experiment which may also show differences at the mean RT level.

(ii) A model-based analysis is a first step in moving beyond the crucial, foundational taxonomy-building stage exemplified by Pomerantz and colleagues (Pomerantz, 1983; Pomerantz & Portillo, 2011; Treisman & Paterson, 1984) to pin down not only whether certain configural features exist, but *how* they are processed, at an algorithmic level. The capacity coefficient allows us to pose questions about the manner in which different sources of information are integrated (or not) in more complex stimuli, about which channels of information are salient in the first place, and about various ways in which processing differs from baseline models of theoretical interest.

(iii) The capacity coefficient provides a more theoretically principled, robust, and interpretable measure of efficiency than mean RT or accuracy can capture. In other words, if we would like to characterize the efficiency with which the perceptual system processes configural features, compared to local features, traditional measures like mean RT and accuracy are often insufficient for discriminating among even basic properties of perceptual processes (e.g., see Townsend, 1990a & Townsend, 1990b).

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 feature discrimination (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, Townsend and Pomerantz (2008). In their study, participants were presented with stimuli akin to those used by Pomerantz, Sager and Stoever (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 analyses of 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 (additionally, these researchers were not ultimately 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. Based on the capacity coefficient predictions, 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.

1.1. 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. Luck & Vogel, 1997; Nosofsky, 1986; Treisman & Gelade, 1980; 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, Mundy, & Tjan, 2012).

[1] See Townsend and Nozawa (1995) and Houpt and Townsend (2012) for mathematical derivation and treatment of the capacity coefficient.

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019

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, which may be processed without decomposition into a more fundamental set of independent features (although such late-stage decomposition may occur on an 'as-needed' basis; see General Discussion). 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 higher-order feature detector or unitization process (Blaha, Busey, & Townsend, 2009; Hendrickson & Goldstone, 2009).

The primordial examples of emergent features arose in the context of grouping. For instance, 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 subsumed by their overall organization, and the phenomenology is controlled by a small set of parameters (Kubovy & Gepshtein, 2003).

Note that this distinction between local and Gestalt theories of features operates on a process- or computational-level of analysis and does not necessarily map onto any clean distinction between regions of neural processing. It may be tempting to associate 'local' features with the properties detected by neurons in low-level visual cortex (e.g. V1) and 'Gestalt' features with properties detected in higher-level ventral stream areas, (e.g. the fusiform face area), but this prediction certainly does not follow from the literature on Gestalt processing. Indeed, there is also evidence that some emergent features, like Orientation and Proximity, may be detected in low-level visual areas (Von der Heydt, Peterhans, & Baumgartner, 1984). Though important, these questions are outside the scope of the general information-processing paradigm that we take in this paper.

1.2. 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 Fig. 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, Sager & Stoever, 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 Fig. 1(a)). In other trials, some non-informative context (Fig. 1(b)) was added to all quadrants to form a composite stimulus (Fig. 1(c)).

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; Pomerantz & Portillo, 2011; Treisman & Paterson, 1984) have postulated a number of emergent features for lines and dots which could account for these results.

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. 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).

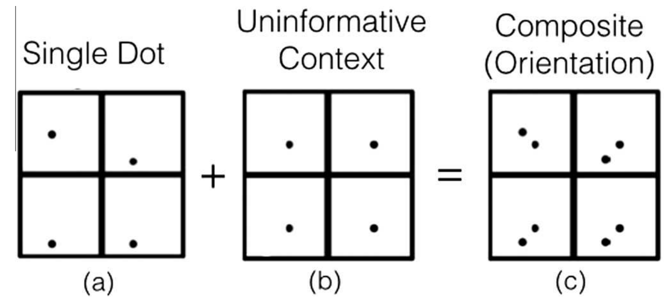

Fig. 1. Example odd-quadrant stimuli adapted with permission from Pomerantz and Portillo (2011). (a) In the single-dot condition, participants were asked to select the quadrant that was different from the others. In this case, the correct response is the upper-left panel. (b) An uninformative context 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).

Pomerantz and Portillo (2011) lay crucial groundwork for building a taxonomy of emergent features, by comparing response times across various conditions. In the present work, we take one step further, investigating not just *what kinds* of emergent features exist, but *how* they are processed at the algorithmic level of analysis. We focus specifically on the simplest case in which emergent features can become salient in visual perception: a pair of dots. In the next section, we motivate our modeling framework, define the capacity coefficient within this framework, and argue that the capacity coefficient confers several unique and novel benefits over traditional measures.

1.3. Systems Factorial Technology

The capacity coefficient is a key component of the modeling framework known as Systems Factorial Technology (SFT; Algom et al. (2015); Houpt & Townsend, 2012; Townsend & Nozawa, 1995; Townsend & Wenger, 2004b; Wenger & Townsend, 2001). SFT provides a set of tools for rigorously defining and testing concepts in the broader information-processing paradigm commonly evoked in cognitive psychology. By abstracting sources of information to 'channels' in an abstract information-processing system, we can rigorously pose a number of algorithmic-level questions about the way our visual system processes various sources of information. For example, in the present work, we ask how the efficiency of processing the whole stimulus changes as parts are added in different configurations. Due to this 'channel' abstraction, we can rigorously define 'efficiency' in terms of stochastic processes in a multi-channel information processor.

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 channel (here, each feature) is processed independently and in parallel. Suppose, in the context of our task, that there is a *left channel L* and a *right channel R*. We can estimate the cumulative hazard function $H(t)$ – the integral over time of the likelihood of the response process terminating at time t given that it has persisted until that point in time – for each channel by collecting response time distributions for a channel in isolation. These two hazard functions are denoted $H_L(t)$ and $H_R(t)$ for the left and right channels, respectively. The parallel race model predicts that if targets are present in both channels (i.e. in a *redundant-target* condition, denoted *LR*) and the participant is asked to respond as soon as a target is observed in *either* channel (i.e. an OR stopping rule), the pertinent cumulative hazard function, $H_{LR}(t)$, should be the sum of the individual channels' cumulative hazard functions. In other words, the ratio of the redundant-target hazard function and the sum of the individual channel hazard functions is equal to one. The capacity coefficient is therefore defined as the ratio:

$$C(t) = \frac{H_{LR}(t)}{H_L(t) + H_R(t)} \tag{2}$$

where, again, H_{LR} is the cumulative hazard function derived from the response time distribution when both sources of information indicate a target simultaneously (i.e. on redundant-target trials) and H_L, H_R are the cumulative hazard functions derived from the response time distribution when each target is presented in isolation. The hazard function 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. Note that these functions utilize the entire RT distribution, licensing stronger inferences than summary statistics like the mean (Townsend, 1990b).

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, respectively. Here we use it instead as a sensitive *relative* measure across conditions. Following (Houpt & Townsend, 2012) we use a *z*-score capacity measure, *Cz*, which is a convenient summary statistic for $C(t)$. This measure focuses on correct response times, although it treats incorrect responses as censoring events for the correct response process (see Houpt & Townsend, 2012, for more details). Because $C(t)$ and the capacity *z* score are different transformations of the same data, we use the terms interchangeably in the text.

In addition to its explicit connection to process-level models of cognition, this formulation of efficiency has several advantages over other measures that could be used, like mean response time or accuracy. First, because there is a clearly defined baseline in terms of an information-processing model, we can interpret the absolute value of $C(t)$ in a meaningful way, unlike mean RT, which is solely used as a relative measure to show a difference between conditions. Second, the capacity coefficient provides a unified space to compare diverse phenomena in vision science (Townsend & Eidels, 2011). Different tasks, different stimuli, or different conditions of the same task may have intrinsically different response demands (e.g. base times), leading to ostensibly different mean RT or accuracy measures. To measure the *efficiency* of processing multiple sources of information together across theses cases requires that variation to be appropriately accounted for. The capacity coefficient achieves this goal by defining as the ratio between multiple channels and single channels.

Finally, although mean RT and accuracy results are sometimes the same as capacity results, they do not license the same inferences. Mean RT and accuracy measures of the configural stimulus do not account for the processing time of individual channels. Thus, comparing mean RTs and accuracies for two-dot displays may be misleading in a redundant-target paradigm. For instance, suppose the configuration in the two-dot configural 'Orientation' condition had a faster mean RT than the configuration in the corresponding control condition, and one used this fact to conclude that Gestalt processing was involved. This conclusion could be flawed: suppose the single-dot components of the configural condition were processed more quickly than the single-dot components of the control condition. Then the faster mean RT in the configural condition could simply be attributed to faster processing in the individual channels without any real gains in *efficiency*. The capacity coefficient would not make this error. It is able to normalize the reaction times for the whole by the reaction times of the parts in order to facilitate this comparison. We attempted to be careful in our experimental design to equilibrate all single-dot trials, but this cannot be expected in general.

For the above reasons, we consider the capacity coefficient to be the primary dependent variable of interest, and perhaps the most valid one. Because of its unprecedented application in this setting, however, we also decided to include results for mean reaction time and accuracy against which the capacity coefficient can be compared. For some tests, all three measures agree, while for others their assessments diverge. We will discuss these points of divergence below, but from the theoretical perspective articulated here, the capacity coefficient takes precedence.

2. Overview of the experiments

Our definition of the capacity coefficient suggests a corresponding experimental paradigm to test the local and Gestalt theories of features in pairs of dots. We set channels *L* and *R* to be the dot on the left and right side of the display, respectively. We thus generated some trials in which participants provide responses for these dots in isolation, to estimate $H_L(t)$ and $H_R(t)$, and other trials in which both dots are present (called 'redundant-target' trials), to estimate $H_{LR}(t)$. To test the local theory against the Gestalt theory, we also designed one condition in which emergent features are present in the redundant-target stimulus and a control condition in which they were not.

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 Fig. 2(a)). 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 (Fig. 2(b) and (c)). The masking duration 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 two main advantages that a redundant-target task holds over the odd-quadrant task introduced by Pomerantz, Sager and Stoever (1977). First, 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. While an interesting phenomenon in its own right, this effect interferes with the lower-level grouping phenomena under investigation. For instance, in Fig. 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 avoids false pop-out effects

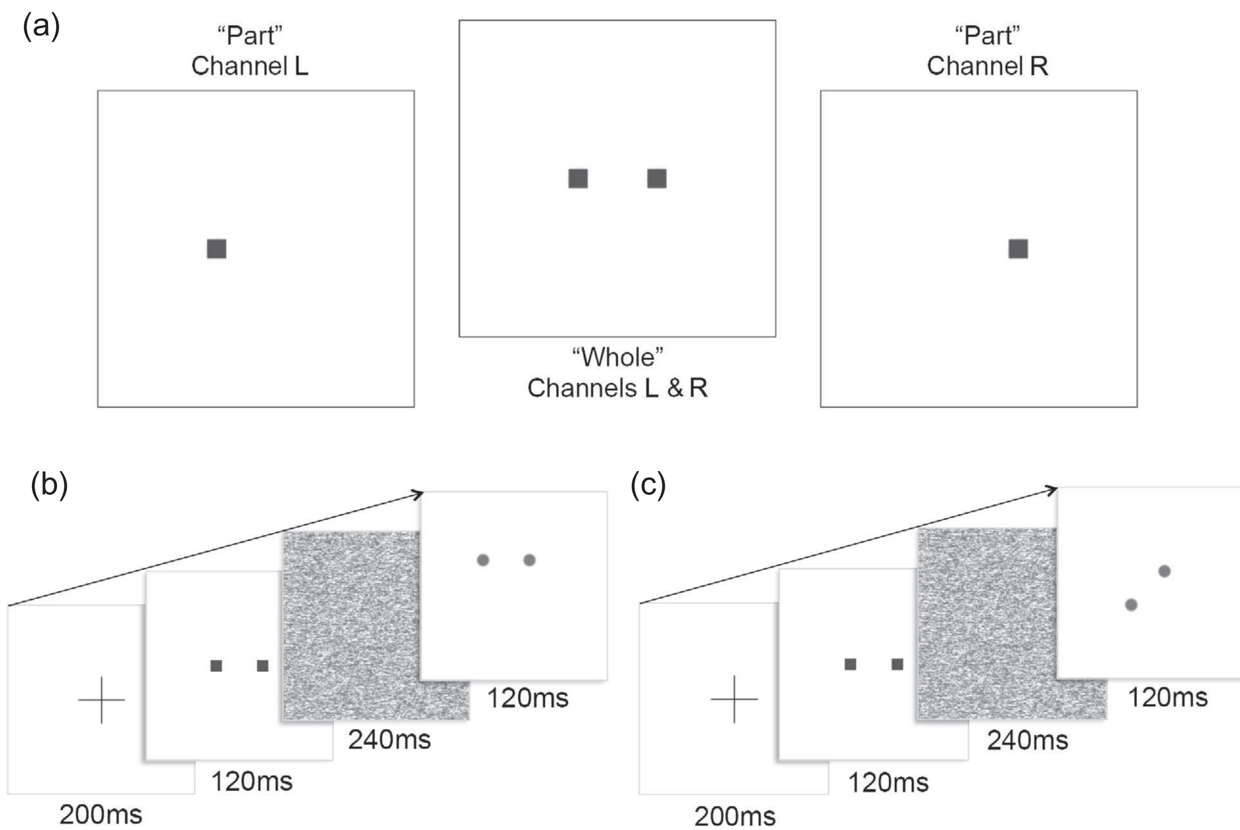

Fig. 2. Stimuli and procedure in change detection task. (a) The three classes of reference stimuli, containing one or both of the channels of local information. (b) The sequence of displays in a 'control' trial. Because the dots changed location in the second frame, the participant should respond 'change'. (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.

by limiting the presentation to a single component or configuration on the screen at a time. Second, the design lends itself to analyses of data using Systems Factorial Technology and its associated measures of capacity.

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 2×2 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. For readers familiar with SFT, 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 modification 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 Fig. 2(b)) or moved in opposite direction to induce a change in emergent feature ("configural"; Fig. 2(c)). 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 a strong 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 that there will not be a strong effect of the line, since the physical features provided by the line were already present as emergent features in the dots. To our knowledge, this is the first study to test this physical vs. emergent feature difference in simple dot stimuli. In the domain of illusory contours, where the Gestalt view of features is well-established, visual discrimination experiments comparing processing of illusory contours vs. real contours found minor speed-ups in reaction time for real contours (Larsson et al., 1999). Since our stimuli are much simpler, if the Gestalt view is correct, any effects of the line in our paradigm would be weak at most. 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.

3. Experiment 1

3.1. Methods

3.1.1. 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 min sessions. Six participants were removed from the study after their first session due to high error rates ($> 30\%$). We pre-set this exclusion criterion based on previous work showing that the $C(t)$ measure is stable up to error rates of approximately 30% and can become unreliable at higher values (Townsend & Wenger, 2004b). 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.

3.1.2. Materials

All stimuli were created using a scripting language for the open-source graphics editor GIMP (Peck, 2006) and presented using the display system DMDX to collect response times (Forster & Forster, 2003) on a 17″ in. ViewSonic CRT monitor (ViewSonic Corporation, Walnut, CA) at 1024×768 resolution with a 75 GHz refresh rate and luminance of 150 cd/m^2. 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 PCI-DIO24 Interface Card (Measurement Computing Corporation, Norton, MA).

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. Fig. 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 (Fig. 3). The implicit line between them is 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 Fig. S2).

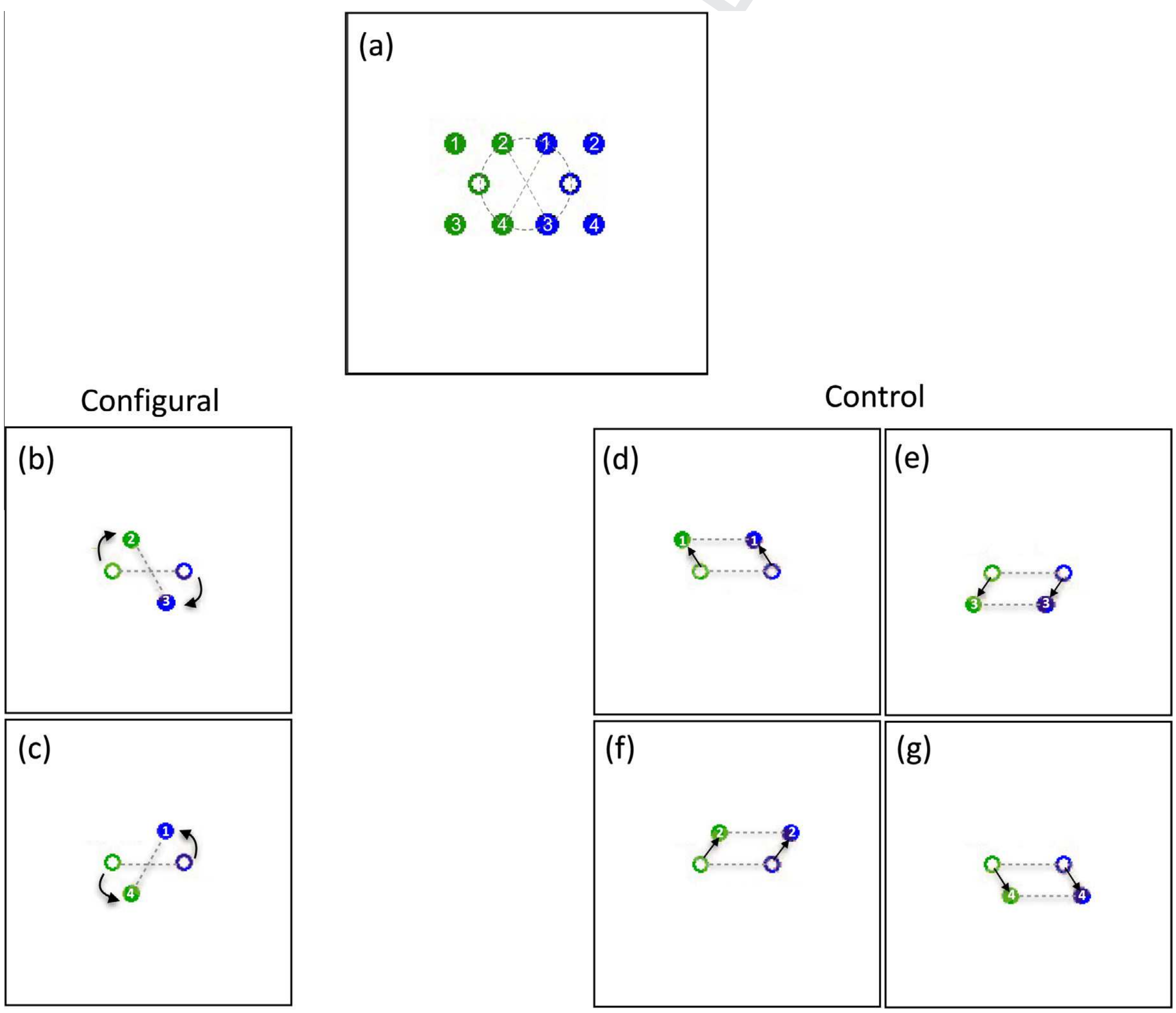

Fig. 3. (a) Possible locations of dots in Experiments 1 and 2. Note that all possible locations for each dot are the same distance away from the reference location, forming an equivalence class under the metric of Euclidean distance. Single-dot stimuli were presented for every position. (b) and (c) Configural 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.

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019

Fig. 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.

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.; panels (d)–(g)) 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.

3.1.3. Procedure

The sequence of displays in a trial is shown in Fig. 2(b) and (c). On each trial, a fixation cross appeared in the center of the screen for 200 ms, followed by a blank display for 27 ms. 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. On line-present trials, the connecting line was only present in the probe, not on the reference screen. The reference screen remained for 120 ms and was then masked for 240 ms by one of five randomly generated Gaussian noise patterns. The probe stimulus was displayed for 120 ms, followed by a blank screen for 1880 ms. 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-min sessions of 960 trials per session. One session contained exclusively ‘line-present’ trials, while the other contained exclusively ‘line-absent’ 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 ‘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 within each block was 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 change (positive response) trials evenly spread over all possible locations. The three varieties of ‘no change’ trials, the two variations of configural trials, and the four variations of control trials were evenly distributed within their respective blocks.

3.2. Results

Bayesian ANOVAs (Rouder et al., 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 Jeffreys, 1961). BF < 3 is weak evidence, and BF < 1 is ‘negative’ evidence, in favor of the null model. Fig. 4 shows the mean response times (a) and accuracies (b) for trials in which two dots were present along. Error bars indicate 95% highest density intervals (HDIs) of the posterior distribution representing our beliefs about the true value of these measures after observing the data. The HDI is the smallest interval of the posterior distribution containing 95% of the density.

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

For capacity we use the (Houpt & Townsend (2012)) *z* score (denoted Cz) 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 Fig. 4(c)) indicated that the most likely model includes a main effect for only configuration ($\mathrm{BF} = 1.2 \cdot 10^{6}$ over a subject only model). Evidence against including an additional main effect of the line was again weak ($\mathrm{BF} = 0.34$) and there was substantial evidence of the configural main effect only model relative to the model with both main effects and an interaction ($\mathrm{BF} = 5.4$). The mean posterior advantage of configural over control on the capacity *z*-scores was 3.15 ($\mathrm{HDI} = [2.14, 4.12]$). The mean posterior difference between capacity *z*-scores without lines and with lines was -0.43 ($\mathrm{HDI} = [-1.29, 0.47]$).

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019

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.

P	Configural						Control						Single dot	
	Lines			No lines			Lines			No lines				
	Z	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	Acc	$\overline{RT}$
1	−1.60	1.00	317	2.87	1.00	349	−3.32	1.00	345	−1.87	0.98	447	0.96	404
2	−4.42	0.99	341	−2.50	1.00	376	−6.74	0.92	430	−6.67	1.00	413	0.98	394
3	−2.32	1.00	563	−4.25	1.00	605	−6.71	1.00	661	−6.42	0.98	620	0.98	641
4	−2.01	1.00	434	−1.58	1.00	354	−3.86	0.94	602	−4.17	0.97	381	0.92	471
5	−0.37	1.00	507	3.99	1.00	404	−6.50	0.98	524	−4.54	0.98	503	0.91	542
6	1.57	1.00	433	0.95	1.00	306	−2.61	0.98	539	−4.70	1.00	377	0.99	450
7	−5.01	0.99	392	−5.11	0.98	331	−6.03	0.94	363	−4.47	0.95	334	0.93	360
8	−2.29	0.77	560	−7.22	0.99	340	−4.67	0.95	451	−5.13	0.98	424	0.89	433
9	−2.79	0.84	433	−5.07	1.00	480	−6.54	0.90	512	−4.02	0.88	577	0.84	547
10	−2.10	1.00	503	−3.22	1.00	509	−6.69	0.99	641	−7.24	0.99	508	0.98	560
11	−4.19	0.97	531	−4.62	0.99	447	−6.35	1.00	541	−4.69	0.99	451	1.00	520
12	−3.39	1.00	351	−0.87	0.99	441	−4.05	0.95	438	−2.59	0.98	417	0.96	430
13	2.88	1.00	511	4.12	1.00	417	−5.56	1.00	469	−3.58	1.00	485	1.00	480
14	−1.53	1.00	429	−0.41	1.00	522	−8.59	0.98	559	−4.43	0.98	460	0.97	512
15	−2.31	0.99	600	−4.72	0.99	561	−6.88	0.86	765	−6.91	0.87	758	0.80	774

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).

3.3. 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 z-score capacity coefficient Cz. 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 Cz. 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 negative direction; response times tended to be 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, with the favored model containing only a main effect of configuration. Any effect of the lines is minor at most. This is a case where accuracy and mean RT point toward a slightly different conclusion than the capacity coefficient, and for the reasons given in the Introduction, demonstrates the advantages of using the capacity coefficient. One explanation of the accuracy and mean RT results 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.

We can also supplement our main analysis by applying the logic of the capacity coefficient to error rates instead of response time distributions. To that end, we compute a summary statistic for the recently developed measure of "accuracy capacity". In particular, we compares accuracy on single-dot trials with the expected accuracy on redundant-target trials under the benchmark race model (cf. Townsend & Altieri, 2012). This summary statistic is given by

$$Cp_{miss} = p(miss|L) \times p(miss|R) - p(miss|LR)$$

where $p(miss|LR)$ is the probability of an error response (i.e., missing a target) in the double target condition, and $p(miss|L)$ and $p(miss|R)$ are the probabilities of missing the targets of single-target trials. Cp is equal to 0 for a baseline parallel, unlimited capacity race model. In Fig. 5, we show this measure plotted on one axis with the

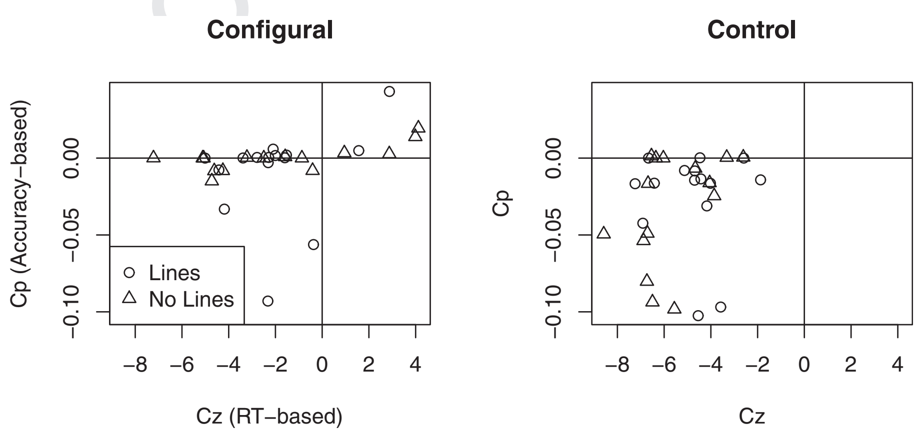

Fig. 5. Experiment 1 scatter plot comparing the z-score capacity coefficient Cz on the x-axis with the accuracy-based capacity assessment function on the y-axis. While the accuracy measure does not yet have formal statistical tests worked out, we can qualitatively see that points in the configural condition tend to be higher on both dimensions than in the control condition.

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019

capacity coefficient z-score on the other axis. We see that both measures tend to be higher in the configural condition than the control condition, giving qualitative evidence that emergent features are processed more efficiently, even when using accuracy as the variable of interest.

Still, it is possible that our 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. To address this concern and also to provide a replication of our results, we ran a second experiment where everything was the same except the blocks were mixed together. This block structure does not allow participants to use different processing strategies a priori.

4. Experiment 2

4.1. Methods

4.1.1. Participants

Twenty paid individuals between the ages of 18 and 26 were recruited from the Indiana University community to participate in two 60 min 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.

4.1.2. Materials

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

4.1.3. 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-min session, we used 1152 trials per 60-min session. Again, one session contained only 'line' trials and the other contained only 'no line' trials, and the distribution of trial types was the same except the 25% dedicated to double-dot change trials was evenly split between 'configural' and 'control' trials.

4.2. Results

Fig. 6 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 ($\mathrm{BF} = 2.7 \cdot 10^{69}$) and of lines ($\mathrm{BF} = 2.3 \cdot 10^{3}$) and was nearly equivocal with respect the presence of an interaction ($\mathrm{BF} = .51$). In the accuracy data, there was decisive evidence for an interaction between the configuration and the presence of lines ($\mathrm{BF} = 104$). When the interaction was disregarded, there was decisive evidence of a main effect of configuration ($\mathrm{BF} = 2.0 \cdot 10^{65}$) and nearly equivocal evidence against a main effect of lines ($\mathrm{BF} = .53$).

Capacity Z scores were again calculated following (Houpt & Townsend, 2012) for each participant in each condition and are shown in Fig. 6(c). Those values were then compared using a Bayesian ANOVA across the configurality-control manipulation and the implicit-explicit line manipulation. The most likely model included only a main effect of configuration ($\mathrm{BF} = 6.8 \cdot 10^{12}$ over a subject only model) however there was only weak evidence for leaving out an additional main effect of the line ($\mathrm{BF} = 2.8$). The analysis did indicate substantial evidence for the configuration only model when compared to a model including both lines and an interaction ($\mathrm{BF} = 8.0$). The mean posterior advantage of configural over control on the capacity z-scores was 5.83 ($\mathrm{HDI} = [4.77, 6.91]$). The mean posterior difference between capacity z-scores without lines and with lines was -0.387 ($\mathrm{HDI} = [-1.32, 0.567]$).

The grand mean for the capacity z scores at the group level was negative, -4.94 ($\mathrm{HDI} = [-5.96, -3.94]$), 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 ($z \approx 0$; see Table 1) (see Table 2).

4.3. 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

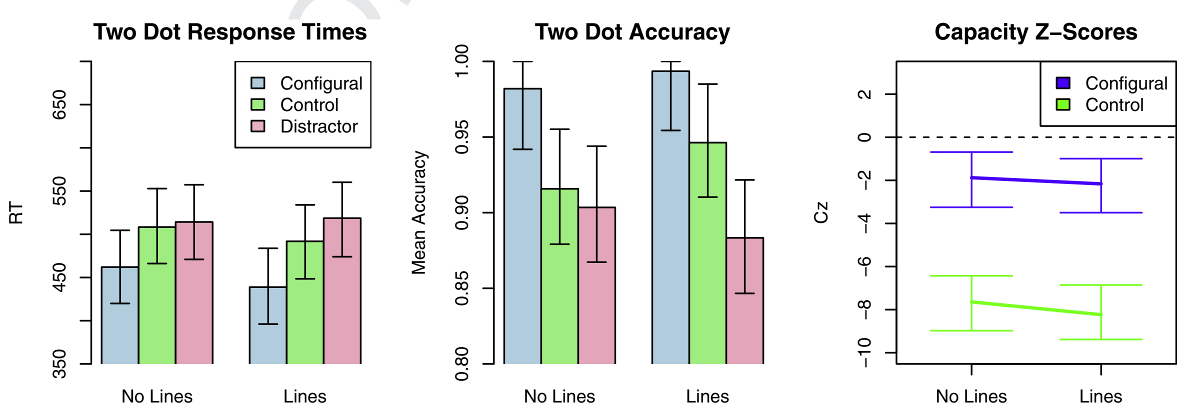

Fig. 6. (a) Mean response times and (b) accuracy for each condition in Experiment 2 (Orientation with mixed configural and control blocks). Configural trials differed from the reference in Orientation as well as the location of each element. In control trials, both elements were in a different location 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.

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019

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

P	Configural						Control						Single dot	
	Lines			No lines			Lines			No lines				
	Z	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	Acc	$\overline{RT}$
1	−3.32	0.99	522	−6.24	1.00	398	−10.76	0.88	608	−7.29	0.90	469	0.88	508
2	−7.55	0.85	549	−8.65	0.99	464	−10.76	0.77	480	−10.19	0.99	493	0.90	454
3	−1.71	0.98	608	−0.19	1.00	371	−10.19	0.96	628	−9.33	0.94	414	0.95	520
4	2.50	0.99	506	2.76	0.99	476	−5.11	0.72	618	−4.21	0.90	534	0.76	546
5	−3.23	0.99	429	−3.52	1.00	334	−7.69	0.97	429	−5.49	0.99	353	0.99	411
6	−4.05	0.99	380	−2.47	1.00	385	−7.63	0.92	444	−3.53	0.95	432	0.93	421
7	−1.91	1.00	456	−2.45	0.97	526	−8.80	0.96	580	−9.53	0.90	637	0.95	607
8	0.11	1.00	409	2.42	1.00	443	−6.93	0.94	483	−8.89	0.92	497	0.93	477
9	−1.12	0.99	401	−0.98	0.99	506	−9.27	0.95	413	−7.74	0.96	528	0.92	476
10	−4.74	0.99	502	−4.23	0.99	471	−7.21	0.94	564	−6.44	0.94	567	0.97	570
11	−0.50	0.98	406	−0.11	0.99	323	−9.96	0.85	506	−10.16	0.97	367	0.90	405
12	−0.18	0.99	543	−1.43	1.00	615	−5.99	1.00	569	−10.81	0.99	698	0.97	644
13	−5.01	1.00	512	0.00	1.00	461	−7.75	0.99	559	−5.05	0.99	484	0.98	543
14	−2.47	0.99	392	−3.39	0.98	466	−7.49	0.90	438	−5.85	0.86	552	0.74	508
15	0.72	0.99	330	0.32	1.00	347	−7.97	0.97	330	−10.12	0.98	371	0.96	363

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 minor 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 measure. It is clear from Fig. 6(b) that the interaction has a fairly small magnitude, and the accuracy measure is not of theoretical interest, 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.

In Fig. 7, we plot "accuracy capacity" alongside participants' capacity coefficient *z* scores, observing that participants tend to be higher on both dimensions in the configural condition. This again reinforces the validity of our measure when participants do not perform at near-ceiling accuracy. 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 simpler 'separated blocks' design from experiment 1.

5. Experiment 3

5.1. Methods

5.1.1. Participants

Twenty-four paid individuals between the ages of 20 and 32 were recruited from the Indiana University community to participate in two 50 min 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

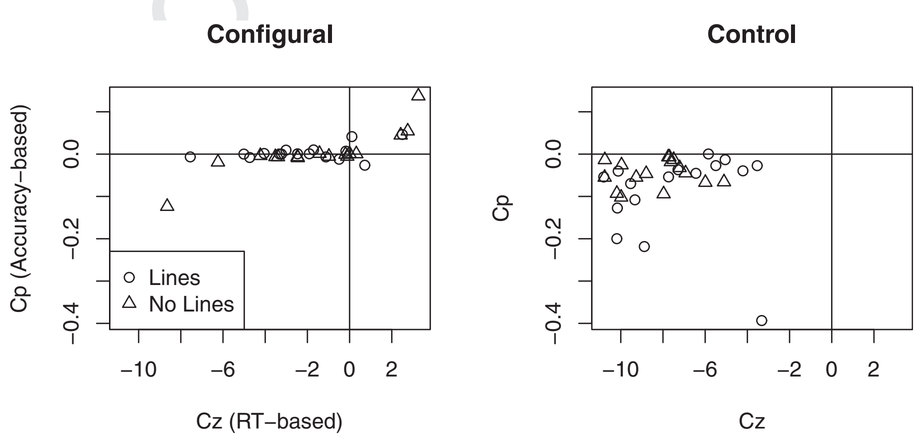

Fig. 7. Experiment 2 scatter plot comparing the *z*-score capacity coefficient on the *x*-axis with the accuracy-based capacity assessment function on the *y*-axis. While the accuracy measure does not yet have formal statistical tests worked out, we can qualitatively see that points in the configural condition tend to be higher on both dimensions than in the control condition.

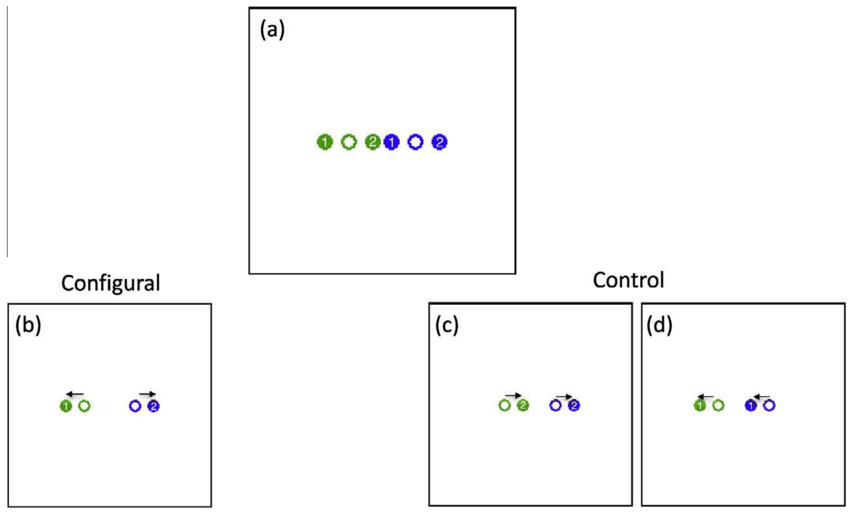

Fig. 8. (a) Possible locations of dots in Experiment 3. Green dots denote possible locations for the left dot, and blue dots denote possible locations for the right dot. Note that all possible locations for each dot are the same distance away from the reference location, forming an equivalence class 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) and (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.

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.

5.1.2. 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. Fig. 8(a) 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; Fig. 8(b)). 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 (Fig. 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 trials, a line connected the two dots.
4. **Control, line present**: Like the other control condition, but a line connected the two dots.

5.1.3. Procedure

The task and protocol were identical to Experiment 1.

5.1.4. Results

Fig. 9 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 ($\text{BF} = 4.6 \cdot 10^{70}$) but very strong evidence against main effect of lines ($\text{BF} = 0.026$) and substantial evidence against a full model including an interaction relative to the model only including a main effect of configuration ($\text{BF} = .11$). 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.[2]

The Bayesian ANOVA on capacity Z scores (shown in Fig. 9(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 ($\text{HDI} = [3.47, 5.41]$). The mean posterior difference between capacity z-scores without lines and with lines was -0.778 ($\text{HDI} = [-1.73, 0.176]$).

Participants were again generally limited capacity, with a group average capacity z-score of -1.56 ($\text{HDI} = [-2.92, -0.0821]$). In one

[2] We ran the analyses including the three low accuracy subjects. The magnitudes of the reported values were slightly different but none of the conclusions changed.

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019

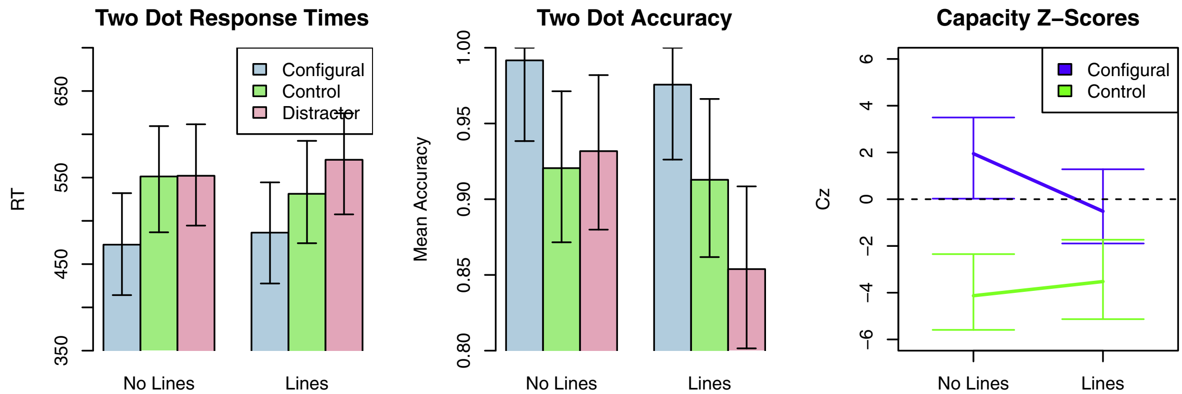

Fig. 9. (a) Mean response times and (b) accuracy for each condition in Experiment 3 (using Proximity and separate configural and control blocks). Configural trials differed from the reference in Proximity as well as the location of each element. In control trials, both elements were in a different location 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.

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

P	Configural						Control						Single dot	
	Lines			No lines			Lines			No lines				
	Z	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	*Z*	Acc	$\overline{RT}$	Acc	$\overline{RT}$
1	−2.29	0.98	438	0.67	0.98	533	−7.67	0.67	573	−5.36	0.66	839	0.76	591
2	0.34	0.98	453	3.56	1.00	511	−3.43	0.80	485	−2.98	0.78	580	0.79	481
3	5.57	0.98	465	2.83	0.98	414	0.02	0.95	531	−5.56	0.95	436	0.89	460
4	−5.82	0.87	494	−1.85	0.99	302	−2.19	0.82	437	−6.53	0.78	375	0.85	413
5	−1.88	1.00	297	2.30	0.99	454	−4.90	0.99	370	−6.39	1.00	583	0.93	509
6	−0.83	0.98	494	0.17	0.99	424	−5.47	0.98	534	−5.92	0.96	505	0.87	506
7	3.97	1.00	420	4.89	0.99	504	0.01	0.97	495	−0.97	0.98	542	0.90	543
8	−4.23	1.00	478	2.31	1.00	540	−8.36	0.98	549	−5.97	0.97	627	0.83	681
9	−3.03	0.99	730	−1.90	1.00	570	−4.32	1.00	630	−6.13	1.00	551	0.92	641
10	2.06	0.99	510	2.45	1.00	342	0.52	1.00	493	−0.39	0.99	458	0.88	538
11	−5.03	1.00	395	−0.33	1.00	461	−5.16	0.99	465	−5.15	0.98	496	0.90	562
12	0.87	1.00	394	2.91	0.98	498	−3.37	0.92	446	−2.46	0.97	532	0.85	512
13	3.60	0.92	778	7.34	0.99	590	−1.47	0.81	960	0.13	0.97	704	0.71	840

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).

5.1.5. 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, the accuracy effect seems to be driven by a larger magnitude drop in the distractor correct rejections between the 'line' and 'no lines' conditions.

In Fig. 10, we plot "accuracy capacity" alongside participants' *z*-score capacity values from Experiment 3 to obtainSupplemental information about configural processing from patterns of errors. We observe that participants tend to be higher on both dimensions in the configural condition, corroborating the statistical tests above.

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. Our Proximity manipulation causes the dots to appear farther apart (due to increased physical distance), while this grouping effect due to connectedness may cause the dots to appear closer together (albeit in psychological distance). This counteracting force would lead to a weaker effect in the 'lines' condition than the 'no lines' condition, where no additional grouping effect was present. 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.

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019

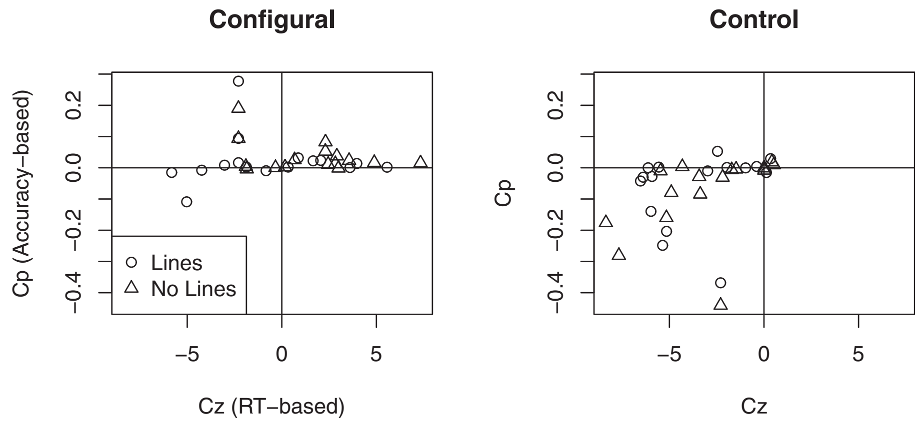

Fig. 10. Experiment 3 scatter plot comparing the capacity coefficient on the *x*-axis with the accuracy-based capacity assessment function on the *y*-axis. While the accuracy measure does not yet have formal statistical tests worked out, we can qualitatively see that points in the configural condition tend to be higher on both dimensions than in the control condition.

Some evidence against this account, however, comes from Han, Humphreys, and Chen (1999), who used a global letter discrimination task to show that grouping elements by Proximity can be as fast and efficient as grouping by connectedness. They found no difference between a condition using only Proximity cues and a condition using both Proximity and connectedness cues. In their experiments, though, Proximity and connectedness were not put in opposition; furthermore, since element connectedness has only been discussed in the context of global grouping tasks, we cannot expect these results to generalize exactly to psychophysical change-detection tasks settings in which only two elements are present. Our task is an example where multiple Gestalt principles come into conflict, which remains an important direction for further investigation.

6. 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 provides a way of rigorously integrating how the parts are processed to make predictions about the whole.

We now turn to some details of our results that raise interesting questions for future work. First, note that while *Cz* 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* ($Cz < 0$), 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 judgement.

If the model containing only local information does not account for the data, we are left with the question of what model *is* appropriate? The SFT framework, and the capacity coefficient in particular, naturally suggests several candidates. These models are unequivocally in the information-processing paradigm, and embody different hypotheses about the sources of information, the order of processing that information, and the way that information is ultimately combined into a decision. All of these aspects of information-processing are intimately tied into the SFT framework and can most easily be framed in terms of its stochastic process-based measures. Further work is needed to distinguish among them, and we suggest some potential variations of our change-detection task that may do so.

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 overpower the channels containing local information in that race. It has recently been suggested that topological similarity may play such a role (Eidels, Townsend & Pomerantz, 2008; Pomerantz, 2003), and is also implicitly endorsed by Pomerantz and Portillo's (Pomerantz & Portillo, 2011) Theory of Basic Gestalts, which posits direct detectors for emergent features. 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 et al., 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

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019

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 (Colonius & Townsend, 1997; Miller, 1982). 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, Townsend & Pomerantz, 2008).

We expect that SFT and the capacity coefficient will be instrumental in distinguishing between these models. SFT was initially developed precisely because of the critical mimicry problems facing traditional measures and analyses. For example, mean reaction time and accuracy measures famously cannot distinguish between parallel and serial architectures in domains like visual search (Townsend & Wenger, 2004). Although it may not be technically impossible to distinguish between the three specific models presented in our General Discussion using traditional measures, we worry about the historical failings of these measures, and expect the tools introduced in this paper to pose fewer problems down the road.

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.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2015.04.019.

References

Algom, D., Eidels, A., Hawkins, R. X., Jefferson, B., & Townsend, J. T. (2015). Features of response times: Identification of cognitive mechanisms through mathematical modeling. *The oxford handbook of computational and mathematical psychology*. Oxford University Press.

Altieri, N., & Townsend, J. T. (2011). An assessment of behavioral dynamic information processing measures in audiovisual speech perception. *Front Psychology, 2*.

Austerweil, J. L., & Griffiths, T. L. (2011). A rational model of the effects of distributional information on feature learning. *Cognitive Psychology, 63*(4), 173–209.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115.

Blaha, L. M. (2011). *A dynamic Hebbian-style model of configural learning* (Unpublished doctoral dissertation), Indiana University.

Blaha, L. M., Busey, T., & Townsend, J. T. (2009). An LDA approach to the neural correlates of configural learning. In *Proceedings of the 31st annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence, 97*(1), 245–271.

Brunelli, R., & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(10), 1042–1052.

Burns, D., Houpt, J., & Townsend, J. (2010). Facial perception as a configural process. *Journal of Vision, 10*(7). 688–688.

Colonius, H., & Townsend, J. T. (1997). Activation-state representation of models for the redundant-signals-effect. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 245–254). Mahwah: NJ Lawrence Erlbaum Associates Publishers.

Eidels, A., Houpt, J. W., Altieri, N., Pei, L., & Townsend, J. T. (2011). Nice guys finish fast and bad guys finish last: Facilitatory vs. inhibitory interaction in parallel systems. *Journal of Mathematical Psychology, 55*(2), 176–190.

Eidels, A., Townsend, J. T., & Pomerantz, J. R. (2008). Where similarity beats redundancy: The importance of context, higher order similarity, and response assignment. *Journal of Experimental Psychology: Human perception and performance, 34*(6), 1441–1463.

Forster, K. I., & Forster, J. C. (2003). Dmdx: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*(1), 116–124.

Garner, W. R. (1974). *The processing of information and structure*. Lawrence Erlbaum.

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychological Science, 23*(4), 427–434.

Han, S., Humphreys, G. W., & Chen, L. (1999). Uniform connectedness and classical gestalt principles of perceptual grouping. *Perception & Psychophysics, 61*(4), 661–674.

Hendrickson, A. T., & Goldstone, R. L. (2009). Perceptual unitization in part-whole judgments. In *Proceedings of the thirty-first annual conference of the cognitive science society* (pp. 1084–1089). Amsterdam, Netherlands: Cognitive Science Society.

Houpt, J. W., & Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology, 56*(5), 341–355.

Houpt, J. W., Townsend, J. T., & Donkin, C. (2014). A new perspective on visual word processing efficiency. *Acta Psychologica, 145*, 118–127.

Jeffreys, H. (1961). *The theory of probability*. Oxford University Press.

Kimchi, R., Behrmann, M., & Olson, C. (2003). *Perceptual organization in vision: Behavioral and neural perspectives*. Taylor & Francis.

Kruschke, J. K. (1992). Alcove: an exemplar-based connectionist model of category learning. *Psychological Review, 99*(1), 22.

Kubovy, M., & Gepshtein, S. (2003). Perceptual grouping in space and in space-time: An exercise in phenomenological psychophysics. In R. Kimchi, M. Behrmann, & C. R. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives* (pp. 45–86). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35*(1), 71–98.

Li, J., & Allinson, N. M. (2008). A comprehensive review of current local features for computer vision. *Neurocomputing, 71*(10), 1771–1787.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology, 14*(2), 247–279.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology: Human Perception and Performance, 17*(1), 160.

Moore, C. M., & Egeth, H. (1998). How does feature-based attention affect visual processing? *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1296.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353–383.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39.

Orsten, K., & Pomerantz, J. R. (2012). False pop out: Evidence of configural disruption in conventional pop out. *Journal of Vision, 12*(9). 1287–1287.

Palmer, S. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review, 1*(1), 29–55.

Peck, A. (2006). *Beginning gimp: From novice to professional*. Apress.

Pinker, S. (1984). Visual cognition: An introduction. *Cognition, 18*(1), 1–63.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112*(4), 516.

Pomerantz, J. R. (2003). Wholes, holes, and basic features in vision. *Trends in Cognitive Sciences, 7*(11), 471–473.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 1–46). New York: John Wiley & Sons.

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., & Portillo, M. (2012). Emergent features, gestalts, and feature integration theory. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with anne treisman* (pp. 187–192). New York, NY: Oxford University Press.

VR 7093
16 May 2015
ARTICLE IN PRESS
No. of Pages 15, Model 5G

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.

Rogosky, B. J., & Goldstone, R. L. (2005). Adaptation of perceptual and semantic features. In L. A. Carlson & E. van der Zee (Eds.), *Functional features in language and space: Insights from perception, categorization and development* (pp. 257–273). England Oxford University Press: Oxford.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default bayes factors for anova designs. *Journal of Mathematical Psychology, 56*(5), 356–374.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*(01), 1–17.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46*(2), 225–245.

Tanaka, J. W., & Farah, M. J. (2003). The holistic representation of faces. In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 53–74). New York: Oxford University Press.

Townsend, J. T. (1990a). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science, 1*(1), 46–54.

Townsend, J. T. (1990b). Truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin, 108*(3), 551.

Townsend, J. T., & Altieri, N. (2012). An accuracy–response time capacity assessment function that measures performance against standard parallel predictions. *Psychological Review, 119*(3), 500.

Townsend, J. T., & Eidels, A. (2011). Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied. *Psychonomic Bulletin & Review, 18*, 659–681. http://dx.doi.org/10.3758/s13423-011-0106-9.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*, 4.

Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review, 11*(3), 391–418.

Townsend, J. T., & Wenger, M. J. (2004b). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review, 111*(4), 1003–1035.

Treisman, A. (1988). Features and objects: The fourteenth bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology, 40*(2), 201–237.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136.

Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance, 10*(1), 12.

Von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*(4654), 1260–1262.

Wenger, M., & Townsend, J. T. (2001). Faces as gestalt stimuli: Process characteristics. In M. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 229–284). Erlbaum Press.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202–238.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*(6), 495–501.

Please cite this article in press as: Hawkins, R. X. D., et al. Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research* (2015), http://dx.doi.org/10.1016/j.visres.2015.04.019