Vision Research 126 (2016) 19–33

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

ARTICLE INFO

Article history:
Received 27 November 2014
Received in revised form 25 April 2015
Available online 15 May 2015

Keywords:
Perceptual organization
Visual perception
Workload capacity

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 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.

☆ *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, Jim Pomerantz and Mary Portillo for illuminating conversations about emergent features and Gestalt theory, 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 Elsevier Ltd. All rights reserved.

The capacity coefficient is a nonparametric measure of workload capacity that derives from an extensive body of work using 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 that 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

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