Vision Research 126 (2016) 291–307

Contents lists available at ScienceDirect

Vision Research

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

Grouping by proximity and the visual impression of approximate number in random dot arrays

Hee Yeon Im[a,c,*], Sheng-hua Zhong[b,c], Justin Halberda[c]

[a] Department of Radiology, Harvard Medical School/Massachusetts General Hospital, United States
[b] College of Computer Science & Software Engineering, Shen Zhen University, China
[c] Department of Psychological and Brain Sciences, Johns Hopkins University, United States

ARTICLE INFO

Article history:
Received 31 January 2015
Received in revised form 21 August 2015
Accepted 21 August 2015
Available online 26 September 2015

Keywords:
Perceptual grouping
Numerosity estimation

ABSTRACT

We address the challenges of how to model human perceptual grouping in random dot arrays and how perceptual grouping affects human number estimation in these arrays. We introduce a modeling approach relying on a modified *k*-means clustering algorithm to formally describe human observers' grouping behavior. We found that a default grouping window size of approximately 4° of visual angle describes human grouping judgments across a range of random dot arrays (i.e., items within 4° are grouped together). This window size was highly consistent across observers and images, and was also stable across stimulus durations, suggesting that the *k*-means model captured a robust signature of perceptual grouping. Further, the *k*-means model outperformed other models (e.g., CODE) at describing human grouping behavior. Next, we found that the more the dots in a display are clustered together, the more human observers tend to underestimate the numerosity of the dots. We demonstrate that this effect is independent of density, and the modified *k*-means model can predict human observers' numerosity judgments and underestimation. Finally, we explored the robustness of the relationship between clustering and dot number underestimation and found that the effects of clustering remain, but are greatly reduced, when participants receive feedback on every trial. Together, this work suggests some promising avenues for formal models of human grouping behavior, and it highlights the importance of a 4° window of perceptual grouping. Lastly, it reveals a robust, somewhat plastic, relationship between perceptual grouping and number estimation.

1. Introduction

We possess the remarkable ability to nonverbally extract the numerosity of collections of multiple items through a near-instantaneous impression of approximate number. This ability can be useful in real world contexts, which often contain structures formed by groups of similar objects clustered together (e.g., trees in a forest or buildings and cars on the street, etc.). In such cases, it is often impractical to directly count items one-by-one: the number of items may be too large, separating already-counted items from not-yet-counted ones may be very difficult, the viewing time may be limited, and so on.

The situations in which we most naturally extract the approximate number of visual elements are also situations that naturally invite perceptual grouping of items into clusters. In the lab, perceived numerosity has often been explored by presenting human observers with simplified images of multiple items that are distributed over space and asking observers to estimate or discriminate numerosity (e.g., Gilmore et al., 2013; Halberda, Sires, & Feigenson, 2006; Izard & Dehaene, 2008; Jevons, 1871; Smets, Gebuis, Defever, & Reynvoet, 2014; Whalen, Gallistel, & Gelman, 1999). Previous findings consistently show that observers can apprehend the approximate number of items from a very brief exposure (e.g., 100 ms without a mask: Izard & Dehaene, 2008; 500 ms of presentation, followed by a mask: Halberda et al., 2006; even from 66 ms of presentation, followed by a mask: Im & Halberda, unpublished data). The rapidity of numerosity estimation seems somewhat surprising given that counting takes about 300 ms per item (Simon & Vaishnavi, 1996). The mechanisms that allow us to quickly and easily perceive the numerosity of up to 100 items within a 100 ms display remain a mystery (e.g., Izard & Dehaene, 2008). The fact that counting requires 300 ms per item also motivates the suggestion that perceived numerosity of a large number of elements may be achieved relying on a distinct

* Corresponding author at: Department of Radiology, Harvard Medical School/Massachusetts General Hospital, United States.

E-mail address: him3@mgh.harvard.ediu (H.Y. Im).

http://dx.doi.org/10.1016/j.visres.2015.08.013
0042-6989/

mechanism from that operating for serial counting of individual elements (for review, see Feigenson, Dehaene, & Spelke, 2004). Furthermore, the ability to extract the approximate number of items in visual collections is present from human infancy (Xu & Spelke, 2000), and is also shared by other animal species (Hauser, Carey, & Hauser, 2000; Meck & Church, 1983). This further suggests that there is a very basic visual mechanism for approximating the number of items in a visual display – a mechanism that does not require schoolroom teaching.

There are several features of numerosity estimation that may help to determine the underlying mechanism. Previous studies of numerosity estimation consistently find that observers underestimate the actual numerosity (e.g., Indow & Ida, 1977; Izard & Dehaene, 2008; Krueger, 1982, 1984). While underestimation is present from the very first trial (Krueger, 1982), Izard and Dehaene (2008) have shown that observers' numerosity estimations can also be calibrated such that observers adjust their estimation either to increase or decrease the amount of underestimation when they are provided with explicit feedback. The source of this underestimation remains to be described, and one possibility is that this underestimation emerges from the heuristic, or algorithm, for extracting approximate number from the visual display.

Human observers' numerosity judgments also display an inherent variability or noise that increases linearly with the signal – scalar variability (discussed as the coefficient of variation, CV: Cordes, Gelman, & Gallistel, 2001; Crollen, Castronovo, & Seron, 2011; Frank, Everett, Fedorenko, & Gibson, 2008; Le Corre & Carey, 2007; or also as the Weber fraction, w: e.g., Dehaene, 2003; Dehaene & Changeux, 1993; Meck & Church, 1983; Stoianov & Zorzi, 2012). CV reflects the normalized standard deviation of assumed Gaussian distributions for internal representations, which is inversely related to the precision of the internal representation. Therefore, the precision of numerosity estimation can be quantified by CV, with lower CV indicating more precise number estimation.

Another feature of numerosity estimation that may inform proposed mechanisms is the lack of a demonstrated upper bound for number estimation. Unlike serial counting of individual objects for small, precise number (e.g., subitizing; Trick & Pylyshyn, 1993, 1994), extracting approximate number is not constrained by the limited capacity of object-based attention. For example, observer's error rate and response time do not increase with the absolute numerosity, suggesting that extracting approximate number does not rely on the serial, limited process of object-based attention (Barth, Kanwisher, & Spelke, 2003). For these reasons, several researchers have suggested that a global process might support the estimation of approximate number, and there are many such global processes that could be relevant. For example, it has been suggested that textural information about the whole scene such as the density of elements within a given area can support the rapid extraction of large, approximate numerosity of elements in a visual array (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Tibber, Greenwood, & Dakin, 2012). Such models are consistent with suggestions that numerosity is not perceived directly, that is, as an independent visual property, but rather is calculated indirectly via texture density (Durgin, 2008). Indeed, one would expect that density and numerosity would be highly inter-related in the environment (e.g., more items goes with more density).

Relatedly, it has been found that how dots are spatially organized can modulate perceived numerosity – e.g., a uniform layout of items throughout the display area results in a scene that appears more numerous than the same number of items clustered into multiple sub-groups (Frith & Frith, 1972); and dots occupying a more extended region of the display area results in a scene that appears to be more numerous than the same number of items clustered into a smaller display region (Bevan, Maier, & Helson, 1963; Binet, 1890; Ponzo, 1928). From results such as these, it seems likely that the mechanisms that support the extraction of approximate number will involve early, rapid, global processing – perhaps with some additional later algorithms that may be attention-dependent.

The models on visual density and texture perception (e.g., Dakin et al., 2011; Tibber et al., 2012) have been popular not only because they are computationally simple and biologically plausible but also because they can very precisely predict human observer's response bias in numerosity estimation. However, the conclusions from these models may mislead one to overlook the fact that human observers are also able to perceive the visual dots in different levels of hierarchy, from individual objects (e.g., how many dots) to configuration of higher-level groups (e.g., how many clusters). Other work suggests that number judgments rely on interactions across multiple levels (e.g., groups and items).

A fourth feature of number estimation is the effect of visual grouping on number judgments. Approximate number estimation is modulated by how elements are grouped and bound together into higher-order objects. The same number of items will look more numerous when regularly arranged than when randomly distributed (Ginsburg, 1976; Taves, 1941), and random patterns look more numerous than clustered patterns (Ginsburg & Goldstein, 1987). The grouping of elements in a display also affects number estimation latencies such that several groups of dots spread out in the periphery of the display are enumerated faster than the same number of dots clustered into one group in the center of the display (van Oeffelen & Vos, 1982), suggesting that parsing of elements into subgroups may occur before enumeration of the elements. Extending these grouping effects into more advanced visual processing, it has also been shown that when higher-order objects are presented (e.g., 3D-like objects consisting of two squares and a connecting line between the squares; Franconeri, Bemis, & Alvarez, 2009; He, Zhang, Zhou, & Chen, 2009), observers tend to more drastically underestimate the number of squares than when the same number of squares are presented as disconnected "lollipops". Note that in such cases the number of elements (e.g., squares and connecting lines), the size of elements, lower-level visual texture cues, and the overall display area, were held constant – suggesting that it is the higher-order grouping cues that drive the effect. These findings together provide evidence that visual grouping cues affect estimation of approximate number, but they do not provide a computational account for grouping and its effects on approximate number.

These features of approximate number estimation can help inform proposals for mechanisms that support the extraction of approximate number. Proposed mechanisms might provide a principled explanation for the underestimation bias, they could explicate proposals that rely on rapid and global processes, and they might include a role for visual grouping effects in numerosity perception. Because perceptual grouping organizes the visual scene into units, and because it can operate rapidly across the entire image, we focus here on the possibility that perceptual groups may be crucial higher-units for the rapid extraction of approximate number in random dot arrays.

Before our empirical investigation, we first consider the literature on perceptual grouping in greater detail. Even when there is no explicit grouping cue such as connecting lines (e.g., Franconeri et al., 2009), the visual system can organize the visual scene easily and flexibly. When similar items are randomly distributed over space, observers can readily and near-instantaneously organize the global structure from the scene by grouping items together based on proximity (Pomerantz, 1981). Visual grouping has been a significant focus of perception research since it was first emphasized by Gestalt psychologists (Wertheimer, 1924). The law of