The Encoding of Spatial Information During Small-Set Enumeration

Harry Haladjian (haladjian@ruccs.rutgers.edu)
Manish Singh (manish@ruccs.rutgers.edu)
Zenon Pylyshyn (zenon@ruccs.rutgers.edu)
Randy Gallistel (galliste@ruccs.rutgers.edu)
Center for Cognitive Science
Rutgers University – New Brunswick
152 Frelinghuysen Road, Piscataway, NJ 08854 USA

Abstract

Using a novel enumeration task, we examined the encoding of spatial information during subitizing. Observers were shown masked presentations of randomly-placed discs on a screen and were required to mark the perceived locations of these discs on a subsequent blank screen. This provided a measure of recall for object locations and an indirect measure of display numerosity. Observers were tested on three stimulus durations (50, 200, 350 ms) and eight numerosities (2-9). Enumeration performance was high for displays containing up to six discs—a higher subitizing range than reported in previous studies. Error in the location data was measured as the distance between corresponding stimulus and response discs. Overall, location errors increased in magnitude with larger numerosities and shorter display durations. When errors were computed as disc distance from display centroid, results suggest a compressed representation by observers. Additionally, enumeration and localization accuracy increased with display regularity.

Keywords: spatial attention; enumeration; subitizing; visual indexing.

I. Introduction

When presented with a set of objects, humans can estimate quickly the set's numerosity with reasonable accuracy. This estimate of number supports various cognitive processes and assists decision-making and action-planning. Given the importance of such abilities, it would be reasonable to expect that a cognitive system employs several methods to obtain numerosity information. The challenge, however, lies in clearly identifying the possible mechanisms involved and determining the conditions under which they are employed.

The primary mechanism responsible for numerosity perception is the nonverbal mental magnitude system that also has been observed in animals and preverbal infants. Magnitudes are inferred mental entities that represent the numerosity or magnitude of things in the world via a mental "accumulator" or "number line" (Dehaene, 1992; Gallistel & Gelman, 1992). An accumulator mechanism is thought to enable the precise representations of duration and numerosity in rats by accumulating neural signals (Meck & Church, 1983). In humans, this accumulator system may represent discrete numerosities through an incrementing process that produces a preverbal count (Gallistel & Gelman, 1992, 2000). Although analog magnitudes are argued to underlie most numerical abilities, an alternate mechanism may be employed for smaller numerosities. The term *subitizing* is used to describe the fast and accurate enumeration of 1-4 objects (Kaufman, Lord, Reese, & Volkmann, 1949). Trick & Pylyshyn (1989, 1994) proposed that a *visual indexing* mechanism may be utilized for subitizing. Visual indexes are "pointers" that automatically pick out and stick to visual items displaying characteristics of "objecthood" (e.g., good continuation, cohesion). Each item that is to be tracked or enumerated is assigned an index in a bottom-up manner, enabling a simultaneous selection of four objects (Pylyshyn, 1989). Subitizing is thought to be the rapid enumeration of these active indexes. When a precise count is required for larger sets, this mechanism can be used to keep track of items that have been counted already, which increases the time required to make a numerosity judgment.

There are theoretical disagreements on the interpretation of the performance differences between small and large sets. Some studies attribute the change in the reaction times to the capacity limitations of information transfer into short-term memory (Cowan, 2001; Klahr, 1973) or a shifting of enumeration strategies (Mandler & Shebo, 1982). The rapid identification of small-set numerosity also can be attributed to the fast mapping of a label to the discrete increments on a mental magnitude (Gallistel & Gelman, 1991) or the fast counting of active indexes (Trick & Pylyshyn, 1994). Whether two systems are responsible for enumeration has yet to be determined conclusively, and this area of research continues to provide evidence supporting both perspectives.

Regardless of the mechanism responsible for subitizing, accurately enumerating a set requires the selection of each visual object. If an indexing mechanism is responsible for subitizing, observers would be able to report on four objects even under time constraints, but with poor memory for locations. Alternatively, if each object must be encoded into working memory for recall, then errors in enumeration and location recall should be similar. Numerosity perception has been studied extensively but little is known about the spatial information that is encoded when enumerating. To address this topic, the current study examines the location encoding that occurs in subitizing.

Studies on the spatial coding of object locations have shown that observers tend to remember locations by using spatial cues to categorize locations according to geometric "prototypes" (Huttenlocher, Hedges, & Duncan, 1991). When presented with a dot inside a geometric shape, children remembered the location as being further away from the midline and edges of that shape—a bias towards the central tendency of the shape category, or prototype (Huttenlocher, Newcombe, & Sandberg, 1994). In adults, the representation of locations also was biased towards the prototype of spatial categories and these biases increased as memory became less certain over extended response delays (Spencer & Hund, 2002). These studies suggest that a single system for representing space is likely to serve both verbal and motor responses that are spatial in nature (Spencer, Simmering, & Schutte, 2006).

One potentially useful approach to understanding enumeration is to apply statistical and computational methods used in the study of visual perception. For example, one recent study used an information theoretic framework to model the human ability to learn statistical regularities from object features in visual displays, and tested whether observers used this information to enhance their ability to identify the locations of specific colors (Brady, Konkle, & Alvarez, 2009). The authors hypothesized that if there were more redundancies in the information input, then more content can be stored (as predicted by information theory). Their results indicate that more regular displays did in fact facilitate the encoding of information, which increased color recall performance in a way that could be predicted by a Bayesian learning model.

The primary goal of the current study is to characterize the spatial encoding during the enumeration of small sets of dots that were randomly placed on a computer screen and to determine if location and enumeration accuracy can be predicted by the statistical or geometric properties of these displays. To investigate this possibility, we devised an enumeration task that presented a display with randomly-placed small black discs. After a mask, observers marked the perceived location of each disc, which also served as their numerosity response (see Figure 1). Three stimulus durations (50, 200, or 350 ms) and eight numerosities (2-9) were tested. These stimuli were presented very briefly in order to prevent verbal counting and the response method allowed for a nonverbal report of numerosity and location (similar to a reporting methodology described in Dent & Smyth, 2006). *Enumeration accuracy* is measured as the percent of trials with an accurate numerosity report and the average (absolute) number of miscounts. For each trial, each disc on a response display was paired with a disc on the stimulus display to determine *location accuracy*, which is the distance between these corresponding discs.

The location data from this experiment was used to characterize observers' representations of objects selected for enumeration. The properties of the disc configurations in the test displays were compared to those in the observers' responses. This enabled quantitative comparisons between the actual stimulus and its representation. One testable prediction is that a display with more regularity would allow more content to be encoded more accurately into working memory, leading to better enumeration performance and object localization. Display regularity was obtained by applying Delaunay Triangulation methods to identify "simplexes"—triangles with vertices comprised of display discs without other discs inside them (Kendall, 1989). This triangulation was applied to the elements in both the test and response displays, and the average area and side lengths of the resulting triangles were computed for each display. "Maximal circles", which connect the vertices of each triangle simplex, have also been used to study regularity in the spacing between dots (Fidopiastis, Hoffman, Prophet, & Singh, 2000). Similarly, maximal circles were identified and the average radii of these circles was computed and compared to observer responses. Another form of statistical summary examined was the centroid of disc configurations. Humans can estimate the center-of-mass of an array of randomly arranged dots on a display with high accuracy (Juni, Singh, & Maloney, 2008; Zhou, Chu, Li, & Zhan, 2006). The computation of this centroid estimate may prove to be crucial when representing individual locations. For each display, we computed the centroid and the distances of each element on the display from its centroid. We then compared the values between the stimulus and response data in order to estimate variability and compression.

The various regularity measures described above may be used to develop a model that predicts enumeration and localization performance. The current study aims to contribute to this goal by characterizing the spatial encoding during enumeration. This can lead to a better understanding of the nature of numerosity representations obtained under brief viewing conditions and help identify the mechanisms that contribute to this process. Using the characteristics of possible mechanisms—such as the Weberian nature of a magnitude mechanism or the set-based limitation of an indexing mechanism—we can test which model best explains the current data and identify the properties that are better predictors of accurate enumeration.

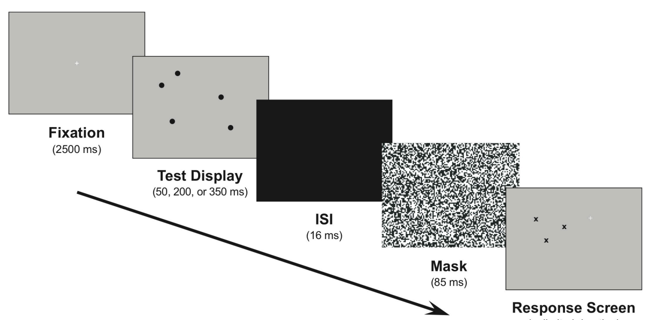

Figure 1. Schematic of this enumeration experiment.

II. Methods

Participants: 24 Rutgers University undergraduates participated in one session for course credit or payment.

Apparatus: The experiment was programmed in MATLAB with Psychophysics Toolbox 3.0.8 (Brainard, 1997) and presented using a desktop computer running Windows XP

(Intel Pentium 4 processor). The stimuli were displayed on a 19" color CRT monitor with a resolution of 1280 x 1024 pixels and a refresh rate of 70 Hz; contrast was set to 100% and brightness was set to 50%. The screen dimensions were approximately 35° by 27° in visual angle.

Stimuli*:* Test displays contained 2-9 identical black discs (35 pixels in diameter, or ~1°) presented on a gray screen for 50, 200, or 350 ms. The discs were randomly placed on the screen with the following constraints: discs could not appear within 115 pixels (~3°) or more than 715 pixels (~20°) of each other, or within ~200 pixels of the screen edges. This produced an effective viewing display of 21° by 16° (768 x 614 pixels). Adequate separation of objects was emphasized to ensure "preattentive" object discriminability, since more attentional resources are required for accurate discrimination when separated by less than 1° (Bahcall & Kowler, 1999). The test display was masked using a random-dot texture created by randomly assigning a white or black value to a grid of 4 x 4 pixel squares.

General procedure: Observers sat approximately 60 cm from a computer screen in a darkened room. They were given instructions by the experimenter and performed six practice trials to ensure understanding of the task. Each trial began with a 2,500 ms presentation of a gray screen with a white central fixation cross. The stimulus screen was then flashed for a designated duration. A black screen appeared for one frame (16 ms) before a mask comprised of a random-dot texture was presented for 85 ms. Finally, a gray input screen with a crosshair pointer appeared and remained until observers made their responses by placing markers ("X") on each of the perceived disc locations. Pressing the space bar initiated the next trial. It was emphasized to the observers that the number of markers placed on the screen should represent the number of discs seen on the test display, even if they were unsure about the exact location. Response coordinates were recorded by the program. See Figure 1 for a diagram of a trial.

Processing the location data: The location data was comprised of two files, one for the stimulus display and another for the response display. In order to analyze the accuracy of location representations, stimulus and response coordinates (x-y values) were paired using the following procedure. When a trial had the same number of stimulus and response elements (i.e., correctly enumerated displays), a Procrustes analysis on the convex hulls of the element locations was used to identify the best fit of the response to the stimulus coordinates for each trial. Procrustes analysis determines the similarity between two shapes by estimating the best fit of one set of points to a comparison set by factoring out variations in scaling, rotation, and translation (Goodall, 1991). After applying the relevant scaling, rotation, or coordinate position transformations, Delaunay Triangulation and nearest-neighbor methods were used to identify stimulus-response pairs. For calculating pattern regularity on a display, the mean and variance values were computed for the areas of triangle simplexes (identified by the triangulation), connecting edges, and the radii of the maximal circles that circumscribe the triangle simplexes. Trials with unpaired discs, which primarily occurred when displays were under- or over-counted, were not included in the location analysis (15% of possible data points).

III. Results

Enumeration Accuracy

The enumeration results replicate previous studies, with the highest accuracy observed in low numerosities. This range was maintained for six items—better than in previous studies where accuracy declines after four items. A follow-up experiment was conducted that included a control where numerosity was reported using Arabic numerals (Haladjian, Pylyshyn, & Gallistel, 2009). Observers performed better in the location-marking block (six items) than the control block (four items), supporting the current results.

Analysis of variance was conducted on the enumeration performance with observer included as a random variable. The largest numerosity condition of nine discs was excluded to control for anchoring effects. Analyzing the proportion of trials with perfect enumeration revealed main effects for display duration (F=34.7(2,276), p<.01) and numerosity (F=68.8(6,276), p<.01), with interactions (F=7.7(12,276), p<.01). Analyzing the absolute value of miscounts for each condition also revealed main effects for display duration (F=36.1(2,276), p<.01) and numerosity (F=51.2(6,276), p<.01), with interactions (F=11.8(12,276), p<.01). Figure 2 depicts the proportion of trials correctly enumerated and Figure 3 depicts the average absolute number of miscounts. Errors increased with larger numerosities but fewer errors were found with longer display durations. When observers made errors, they were generally underestimates (84% of errors were underestimates). Performance in the 50-ms display was significantly worse than the 200- and 350-ms durations for the 6-9 disc displays in both these analyses.

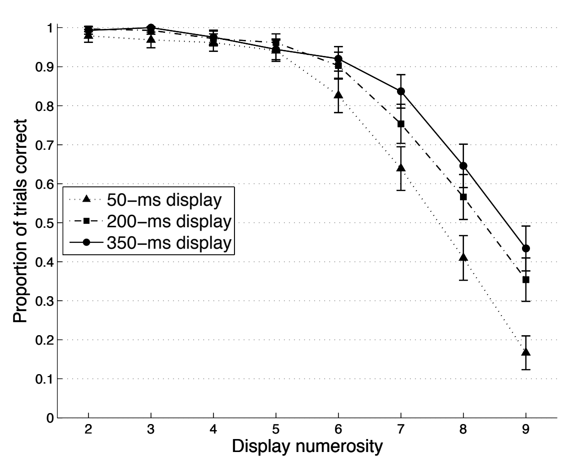

Figure 2. Proportion of trials with correct enumeration.

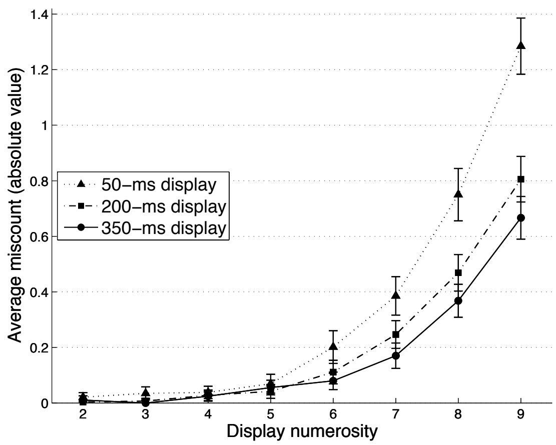

Figure 3. Average counting errors.

Location Accuracy

Location error is reported as the Euclidean distance between the coordinates of stimulus-response pairs for each trial. ANOVA results indicate main effects for display duration (F=27.2(2,276), p<.01) and numerosity (F=81.4(6,276), p<.01), with no interactions (F=1.4(12,276), p=.15). Errors increased with larger numerosities and generally decreased with longer display durations (see Figure 4). The mean and variance of the following variables were computed to estimate display regularity: 1) area of Delaunay "simplex" triangles; 2) length of the triangle segments (shared edges were counted only once); 3) radii of the maximal circles that circumscribed the simplexes; 4) distance between each disc and the display centroid; and 5) radius of the enclosing "circumcircle" around the display elements (to estimate disc dispersion). Since performance was significantly worse in the 50-ms displays, only data from the 200- and 350-ms display durations (combined) are reported here.

The centroid (or center-of-mass) for each display was computed by calculating the mean x- and y-coordinate of all discs on a display. The compression measure is shown in Figure 5 as the average centroid-to-disc distances, that is, the average distance from discs on a display to the centroid. The substantially smaller distances in the observers' responses suggests that their representation is compressed around the centroid of the display. The average dispersion (minimum enclosing circle radius) of the discs on a stimulus display ranged from 203 pixels (SD=73) in 2-numerosity displays to 358 pixels (SD=19) in 9-numerosity displays; for response data, this dispersion ranged from 185 pixels (SD=73) to 314 pixels (SD=44), indicating compression.

Display regularity was measured in terms of the variability in the size of the Delaunay simplexes and the size of the maximal circles that circumscribe these triangles. Here we report the effects of regularity as measured by the variability in the edge lengths of Delaunay simplexes; however, similar patterns of results were obtained with the area of the simplexes and the size of the maximal circles. Figure 6 depicts the average segment lengths and also suggests a compression of these representations. To compare levels of display regularity, the standard deviation of the triangle segments in the test displays were grouped into quartiles, where 25% of the trials with least variation are in the first quartile and 25% of trials with the most variation are in the last quartile. This allowed us to plot location errors as functions of increasing variability (decreasing regularity) in Figure 7 and counting errors in Figure 8. These two charts show that displays with lower variability produce lower errors in both counting and localization (counting performance for displays <6 items are not shown since observers performed almost perfectly).

To compare the regularity of the test and response patterns, the overall compression in the response patterns was first undone using the scaling estimate from the Procrustes analysis. The variance in the simplex segment length for these "uncompressed" response patterns was then compared to, and found to be lower than, the variance in the corresponding stimulus patterns. This suggests that observers imposed regularity on the response patterns than there was not present in the stimulus patterns. Figure 9 plots stimulus and response data from two representative trials, which illustrates the imposed compression and regularity.

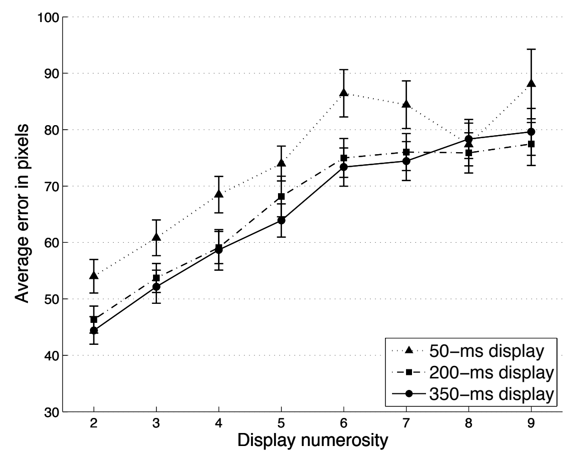

Figure 4. Average location errors in pixels.

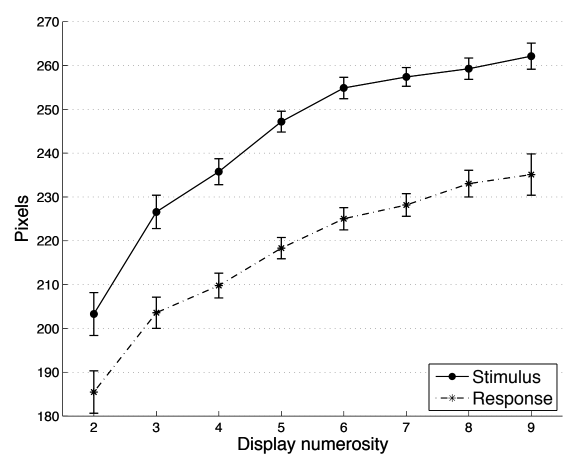

Figure 5. Average centroid-to-disc distance in pixels (200 & 350 ms displays combined).

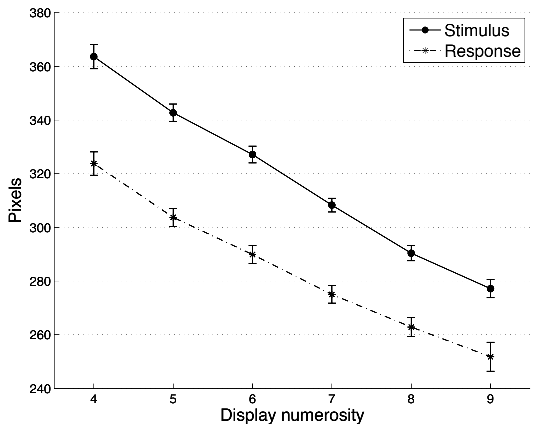

Figure 6. Average segment lengths of Delaunay triangle simplexes (200 & 350-ms displays combined).

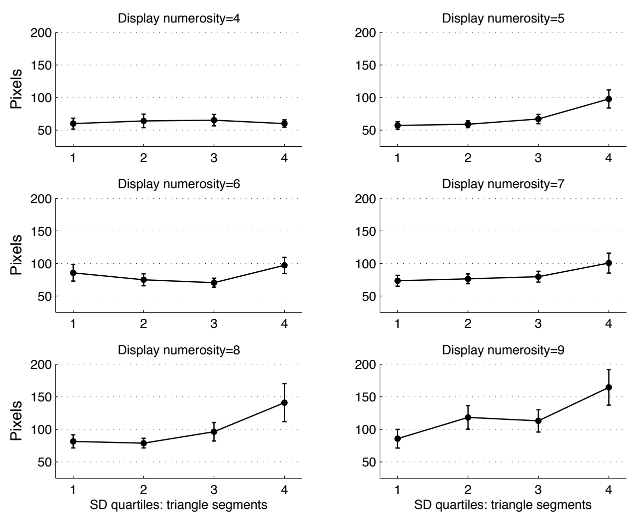

Figure 7. Location errors as a function of increasing triangle segment variability (200 & 350-ms displays combined).

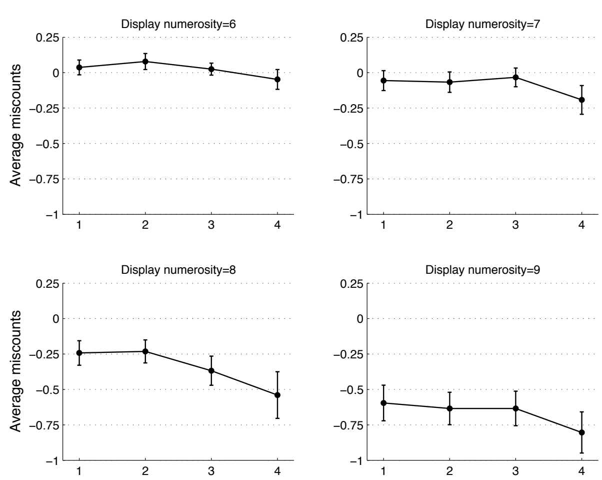

Figure 8. Counting errors as a function of increasing triangle segment variability (200 & 350-ms displays combined).

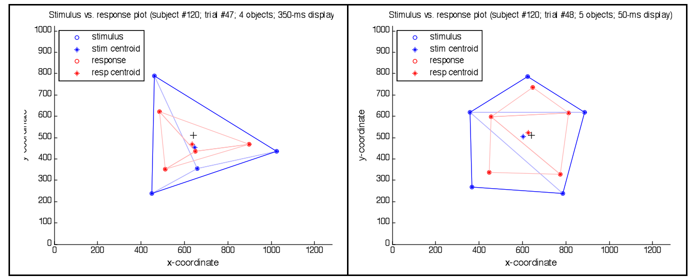

Figure 9. Representative samples of location data with the triangulation simplexes drawn.

IV. Discussion

The visual system is thought to use redundancies from visual stimuli in order to encode information efficiently, as proposed by information theory applications to perception (Attneave, 1954). The current results showing better performance in displays with more regular patterns indicates a more efficient encoding of object locations that may be supported by an information theory of perception. When the triangle simplexes of a display have less variance, observers are more accurate in representing these more regular displays and exhibit better enumerating and localization performance. Additionally, there appears to be a tendency for compressing distances around the centroid. Even after factoring out the overall compression in the response patterns, these distances were found to be less variable in the response configurations than in the test configurations. This could indicate that observers are either assuming there is more regularity when they reconstruct the image, or representation errors are biased towards less variability or towards more "prototypical" representations of shape. This observed tendency to impose regularity on variable displays supports findings from previous studies (e.g., Taylor, 1961).

Increasing stimulus exposure durations from 50 ms to 200 ms produced more accurate enumeration for numerosities greater than six and more accurate location encoding for all numerosities. This suggests a coarse location-estimation process that occurs initially and is updated over time. The disassociation in enumeration and location performance for the smaller numerosity range also suggests that enumeration occurs independent of location-encoding: attention may be required to effectively encode locations but subitizing may be preattentive. This may indicate that visual indexes are responsible for subitizing, since location information does not need to be encoded initially to assign an index, but over time information can be bound to these indexes in order to build more accurate feature representations, including locations (Pylyshyn, 1989). The current results suggest that the indexing mechanism is implemented for smaller numerosities, but further experiments to support this conclusion are required.

The current experiment describes a novel methodology that implements a nonverbal report of numerosity, which appears to enable high enumeration accuracy of six items.

Allowing observers to enumerate by location may be a more accurate demonstration of selection abilities during fast enumeration, and this type of selection is sensitive to the geometric and statistical properties of the visual input. The observed location errors occur systematically and may benefit from inherent geometric regularities. Further analyses of these location data from a statistical perception or information theoretic perspective promise to reveal important information about the spatial nature of numerosity representations.

Acknowledgments

This research was supported by NSF 0549115: IGERT program in perceptual science at Rutgers University (HHH).

References

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*(3), 183-193.

Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research, 39*(1), 71-86.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: Using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General, 138*(4), 487-502.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114; discussion 114-185.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*(1-2), 1-42.

Dent, K., & Smyth, M. M. (2006). Capacity limitations and representational shifts in spatial short-term memory. *Visual Cognition, 13*(5), 529-572.

Fidopiastis, C., Hoffman, D. D., Prophet, W. D., & Singh, M. (2000). Constructing surfaces and contours in displays of color from motion: The role of nearest neighbors and maximal disks. *Perception, 29*(5), 567-580.

Gallistel, C. R., & Gelman, R. (1991). Subitizing: The preverbal counting process. In W. Kessen, A. Ortony & F. I. M. Craik (Eds.), *Memories, thoughts, and emotions: Essays in honor of george mandler* (pp. 65-81). Hillsdale, N.J.: L. Erlbaum Associates.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*(1-2), 43-74.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences, 4*(2), 59-65.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society. Series B (Methodological), 53*(2), 285-339.

Haladjian, H. H., Pylyshyn, Z. W., & Gallistel, C. R. (2009). *Enumerating by location increases the subitizing limit*. Paper presented at the 17th Annual Object Perception, Attention, and Memory (OPAM) Meeting of the Psychonomic Society.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychol Rev, 98*(3), 352-376.

Huttenlocher, J., Newcombe, N., & Sandberg, E. H. (1994). The coding of spatial location in young children. *Cogn Psychol, 27*(2), 115-147.

Juni, M. Z., Singh, M., & Maloney, L. T. (2008). Testing for robustness in visual localization of dot clusters without part structure [abstract]. *Journal of Vision, 8*(6), 1014a.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*(4), 498-525.

Kendall, D. G. (1989). A survey of the statistical theory of shape. *Statistical Science, 4*(2), 87-120.

Klahr, D. (1973). Quantification processes. In W. G. Chase (Ed.), *Visual information processing* (pp. 3-34). New York: Academic Press.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111*(1), 1-22.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9*(3), 320-334.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the finst spatial-index model. *Cognition, 32*(1), 65-97.

Spencer, J. P., & Hund, A. M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology: General, 131*(1), 16-37.

Spencer, J. P., Simmering, V. R., & Schutte, A. R. (2006). Toward a formal theory of flexible spatial behavior: Geometric category biases generalize across pointing and verbal response types. *Journal of Experimental Psychology: Human Perception and Performance, 32*(2), 473-490.

Taylor, M. M. (1961). Effect of anchoring and distance perception on the reproduction of forms. *Perceptual and Motor Skills, 12*, 203-230.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*(1), 80-102.

Zhou, X., Chu, H., Li, X., & Zhan, Y. (2006). Center of mass attracts attention. *Neuroreport, 17*(1), 85-88.