

Enumerating by pointing to locations: a new method for measuring the numerosity of visual object representations

Harry Haroutioun Haladjian · Zenon W. Pylyshyn

Published online: 6 November 2010
© Psychonomic Society, Inc. 2010

Abstract The fast and accurate enumeration of a small set of objects, called *subitizing*, is thought to involve a different mechanism from other numerosity judgments, such as those based on estimation. In this report, we examine the subitizing limit using a novel enumeration task that obtained the perceived locations of enumerated objects. Observers were shown brief masked displays (50, 200, and 350 ms) of 2–9 small black discs randomly placed on a gray screen and then asked to place a marker where each disc had been located. The number of these markers provided an estimate of the number of items processed. This “pointing” methodology enabled observers to accurately “enumerate” displays containing up to six items in contrast with the four-item limit typically found when using standard reporting methods (and replicated here in Experiment 2). These results suggest a different account of the limits found in most subitizing and enumeration studies.

Keywords Attention · Spatial perception · Subitizing · Enumeration

Introduction

The ability to identify and locate a number of objects in a visual scene serves many cognitive functions, such as navigating through crowded environments or playing team sports. Counting is a related ability, where accurate enumeration requires individuating discrete perceptual objects.

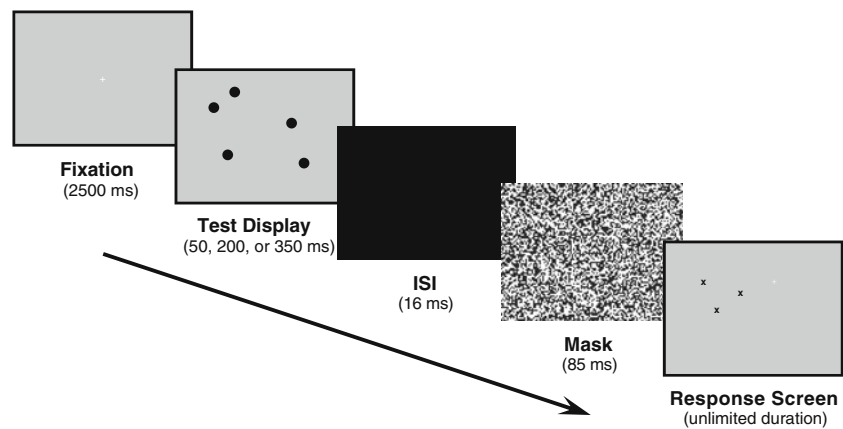
Some theories of enumeration posit that it is achieved via a magnitude estimation mechanism. This system is observed in animals and supports the internal representations of numerosity and duration (Meck & Church, 1983). This mechanism is believed to also enable verbal counting in humans when number labels are mapped to discrete units on this continuous representation (Gallistel & Gelman, 1992).

Results from some studies, however, suggest that different mechanisms may be responsible for computing small numerosities. When enumerating sets of 1–4 items, observers make few errors with modestly increasing reaction times (RT) as set size increases. The rate of errors and RT, however, increases substantially for sets with more than four items (Trick & Pylyshyn, 1993). The term *subitizing* refers to this quick and accurate small-set enumeration (Kaufman, Lord, Reese, & Volkman, 1949). This observation has led some to argue that, in addition to magnitude representations, another system can individuate and select up to four items in parallel and then numerals can be mapped to these items (Feigenson, Dehaene, & Spelke, 2004; Trick & Pylyshyn, 1994a, 1994b). The interpretation of these results, however, remains contested. Some studies attribute the difference in small- and large-set enumeration to capacity limitations of information transfer into short-term memory (Cowan, 2001; Klahr, 1973). Others posit subitizing as pattern-recognition, where familiar patterns formed by fewer items are recognized more quickly, for example, like the patterns on dice (Mandler & Shebo, 1982). Another view argues that the RT slope does not change suddenly but rather increases as a continuous function of increasing variability (Whalen, Gallistel, & Gelman, 1999). Whether or not two separate systems serve enumeration remains an open question.

In this report, we introduce a new experimental methodology designed to examine the relationship between

H. H. Haladjian (✉) · Z. W. Pylyshyn
Department of Psychology, Center for Cognitive Science, Rutgers
University - New Brunswick,
152 Frelinghuysen Road,
Piscataway, NJ 08854-8020, USA
e-mail: haladjian@rucss.rutgers.edu

Fig. 1 Schematic of a trial in this enumeration-by-pointing experiment



spatial representations and the representation of sets of objects in order to characterize the mechanism that supports subitizing. Specifically, we designed an experiment that measures the accuracy of the spatial encoding of objects and indirectly provides an indication of how many objects were recalled, which serves as a measure of enumeration.¹ Observers were shown a brief stimulus (50, 200, or 350 ms) comprised of 2–9 small black discs randomly placed on a gray screen and immediately masked. Then the observers used a mouse to “point to” where the objects had been by placing markers on a blank screen at the former locations of each disc. This methodology allowed us to analyze both location and enumeration accuracy and their relationship.

Experiment 1

Method

Twenty-four Rutgers University undergraduates participated in one 45-min session for course credit or payment. The experiment was programmed in MATLAB[®] using Psychophysics Toolbox 3.0.8 (Brainard, 1997) and controlled by a PC computer running the Windows[®] XP operating system. The stimuli were displayed on a 19-inch (c. 48.3 cm) color HP P1100 CRT monitor (1,280 × 1,024 pixel resolution at 70 Hz).

The test stimulus consisted of 2–9 identical black discs that appeared on a gray background (to reduce contrast and minimize after-images and phosphor decay). These discs were 35 pixels in diameter ($\sim 1^\circ$ visual angle) and randomly placed on the screen with the following constraints: disc edges could not lie within 115 pixels ($\sim 3^\circ$) and no farther than 715 pixels

($\sim 20^\circ$) of each other; additionally, discs could not appear within ~ 200 pixels ($\sim 5^\circ$) from the screen edges. This produced an effective viewing display of 21.1° by 16.6° (768 × 614 pixels). The minimum distance between discs was set at approximately 3° since attention requires at least 1° of visual separation for accurate discrimination (Bahcall & Kowler, 1999) and at least 2° when stimuli extend 15° into the periphery (Intriligator & Cavanagh, 2001).

Observers sat approximately 60 cm from the screen in a darkened room. They were instructed to look carefully for the brief display of black discs in order to notice the number of discs and remember their locations. Each trial began with a 2,500-ms gray screen with a white central fixation cross, on which the observers were instructed to fixate. Then, the test display was flashed for 50, 200, or 350 ms, followed by a 16-ms black screen and an 85-ms mask (created by randomly assigning a white or black value to a grid of 4×4 pixel squares). Finally, a gray screen with a crosshair cursor appeared and observers placed markers (“X”) on the recalled location of each disc (see Fig. 1). It was emphasized that the number of markers placed on the screen should correspond to the number of discs on the test display, even if the observer was unsure about their exact locations. When the observers were finished marking the disc locations, they pressed the space bar to start the next trial. Observers received 12 trials of each of the 24 test conditions (3 durations and 8 numerosities); these 288 trials were randomly distributed throughout the experiment. Observers were encouraged to take a break at any point during the experiment. The primary measures of interest were enumeration accuracy and the magnitude of location errors, which were analyzed using within-subjects ANOVA.

Results

Enumeration accuracy

Numerical accuracy, measured as the proportion of trials in each condition in which the observer provided the correct

¹ Note on terminology: We measure the number of items recalled by counting the number of item-locations observers marked. We call this the number of items “enumerated” (even though the observer does not actually provide a numerical response) because the number of item-locations marked is a measure of the number of items that the observer has attended and recalled and thus in effect non-symbolically enumerated.

number of location marks, was high for displays containing up to six items and decreased significantly for larger numerosities. The ANOVA indicates main effects for display duration [$F(2, 6,336)=85.9, p<0.001, \eta^2=0.789$] and numerosity [$F(7, 6,336)=128.4, p<0.001, \eta^2=0.848$], with an interaction [$F(14, 6,336)=8.0, p<0.001, \eta^2=0.258$]. Performance in the 50-ms display duration was significantly worse than the 200-ms and 350-ms durations for displays with 6–9 items. Figure 2 shows the enumeration accuracy as a function of numerosity for each display duration. (Note: all error bars in this report represent 95% confidence intervals.)

We also analyzed the average number of miscounts in each condition. Over- and under-counting were treated the same in this analysis by taking the absolute value of miscounts (84% of errors were underestimates). ANOVA results for miscounts also indicate main effects for display duration [$F(2, 6,332)=90.6, p<0.001, \eta^2=0.798$] and numerosity [$F(7, 6,332)=92.5, p<0.001, \eta^2=0.801$], with an interaction [$F(14, 6,332)=22.3, p<0.001, \eta^2=0.492$]. The average counting error increased with greater numerosities, but less so for the longer display durations. (See Fig. 3).

Additionally, a simple linear function was computed between display numerosity and response numerosity (combined durations). For small sets (2–6 items), the degree of linear fit between display numerosity and numerical responses was high: adjusted $r^2=0.971$ ($p<0.001$) and $\beta=0.985$. For larger sets (7–9 items), the fit was not as good: adjusted $r^2=0.350$ ($p<0.001$) with a lower slope ($\beta=0.591$).

“Pointing” accuracy

Location error in the pointing task is reported as the Euclidean distance between a stimulus disc and a paired

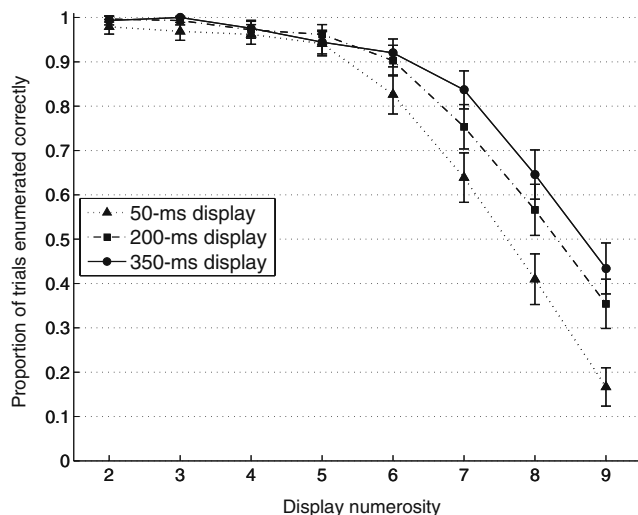


Fig. 2 Proportion of trials enumerated correctly (all error bars represent 95% C.I.)

response disc, which was determined as follows. Using Delaunay Triangulation (Kendall, 1989) and nearest-neighbor methods, we identified the likely associated response marker for each stimulus disc in each trial and then calculated the pixel distance between the centers of these paired discs. Some trials resulted with unpaired discs, for example, when an observer miscounted the stimulus. These trials were excluded from the location analysis (approximately 15% of total trials).

ANOVA results for the magnitude of location errors in each condition indicate main effects for display duration [$F(2, 4,952)=20.7, p<0.001, \eta^2=0.430$] and numerosity [$F(7, 4,952)=66.3, p<0.001, \eta^2=0.819$], but without an interaction [$F(14, 4,952)=1.3, p=0.187, \eta^2=0.051$]. Figure 4 shows the average error distance in pixels and degrees of visual angle. Errors increased for larger numerosities and in the shortest duration—the 50-ms display was significantly worse than the other durations in all numerosities except 8. A regression analysis on the combined durations showed a larger increase (slope) of location errors with numerosity for displays with 2–6 items ($\beta=0.258, \text{adjusted } r^2=0.066, p<0.001$) than for displays with 7–9 items ($\beta=0.182, \text{adjusted } r^2=0.033, p<0.001$).

Experiment 2

The results of Experiment 1 suggest that the “pointing method” allows more items to be processed (and indirectly enumerated) in subitizing than typically reported (e.g., Trick & Pylyshyn, 1994b). Since subitizing experiments have used a variety of methods, there remains the possibility that the subitizing range increase found in this experiment may be due to aspects of our methodology other

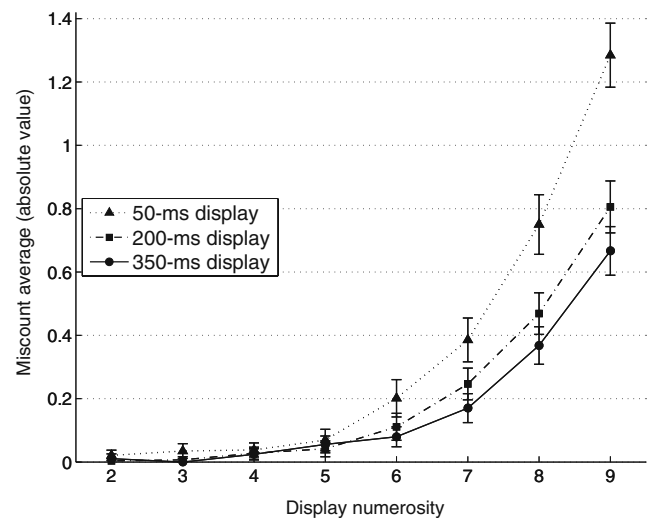


Fig. 3 Average enumeration errors (absolute value)

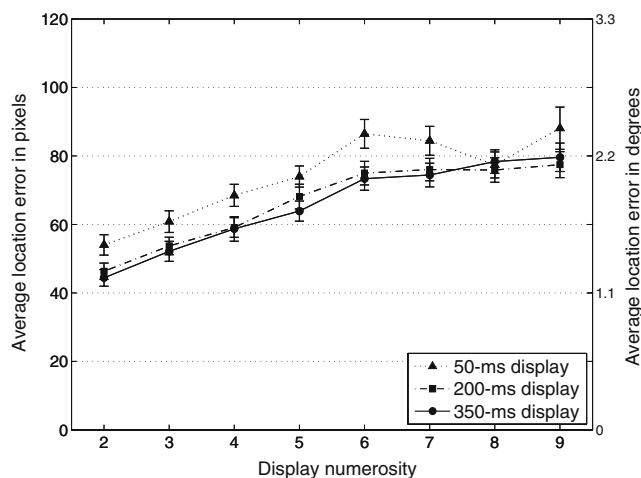


Fig. 4 Average distance between stimulus-response pairs, in pixels (left y-axis) and visual angle in degrees (right y-axis)

than pointing to recalled object locations. Therefore, Experiment 2 compared our indirect “pointing method” of inferring how many items had been processed with the more conventional method that relies on observers’ explicit report of the cardinality of the set. Other aspects of the experiment were the same as in the first experiment (e.g., the nature of the displays, the use of the mouse to report cardinality, performance measures). Since Experiment 2 involves the use of a different set of observers, we replicated our pointing method on the new subject population in order to provide a within-subjects comparison of the pointing response versus the symbolic numeral response.

Method

Nineteen Rutgers University undergraduates participated in one 50-min session for course credit or payment. Experiment 2 consisted of two blocks. In the first (control) block, observers simply reported the number of objects by clicking on the corresponding Arabic numeral on the screen. The 20-point Helvetica font numerals (1–12) appeared on the screen equidistant from the central fixation in the form of a clock-like ring with a radius of $\sim 3.8^\circ$ (140 pixels). The cursor always appeared at the location of the fixation cross and observers clicked on the appropriate number using the mouse pointer. The second block presented the same pointing task described in Experiment 1. This order was always maintained to discourage the use of pointing strategies in the control block (there were no practice effects in Experiment 1). This experiment tested eight numerosities (2–9) and two display durations (50 and 200 ms) for a total of 16 test conditions that were administered 10 times in each block.

Results

Enumeration accuracy

To compare performance between the two reporting methodologies, ANOVAs were performed for the two display durations separately with reporting methodology (explicit numerical response versus pointing response) and numerosity as factors (see Fig. 5). In the 50-ms displays, ANOVA on the proportion of correct trials showed main effects for reporting method [$F(1, 2,768)=20.8, p<0.001, \eta^2=0.536$] and numerosity [$F(7, 2,768)=184.8, p<0.001, \eta^2=0.911$], with an interaction [$F(7, 2,768)=6.7, p<0.001, \eta^2=0.271$]. In the 200-ms displays, there were also main effects for reporting method [$F(1, 2,768)=6.4, p<0.05, \eta^2=0.262$] and numerosity [$F(7, 2,768)=95.5, p<0.001, \eta^2=0.841$], with an interaction [$F(7, 2,768)=5.1, p<0.001, \eta^2=0.220$].

“Pointing” accuracy

ANOVA results for the magnitude of location errors were similar to the results of Experiment 1 and indicate main effects for display duration [$F(1, 2,336)=67.9, p<0.001, \eta^2=0.819$] and numerosity [$F(7, 2,336)=28.1, p<0.001, \eta^2=0.652$], with an interaction [$F(1, 2,336)=2.4, p<0.05, \eta^2=0.138$]. (See Fig. 6).

Experiment 2 shows that the increased subitizing limit observed in Experiment 1 was not due to any incidental properties of the display or the presentation, but can be attributed to the need to respond by pointing to individual items rather than to a symbolic representation of the set’s cardinality.

General discussion

In this study, we explored a novel and indirect way of determining how many briefly-presented items can be individuated and retained for further processing. Observers used the mouse to indicate locations of each item in a set of 2–9 discs that were displayed briefly and masked. By asking observers to indicate *where* each disc had been located, we showed that observers can attend to and recall up to six items. This capacity is in contrast to that obtained when observers only indicated *how many* items there were. The latter limit is generally known as the *subitizing* limit and has been widely reported to be around four items (Trick & Pylyshyn, 1994b).

Performance in reporting locations was also highly accurate (average error distance 2.5°), compared to the mean distance between stimulus objects (over 6°). Location accuracy, however, decreased as the number of objects increased, even for small-set displays with 2–6 objects

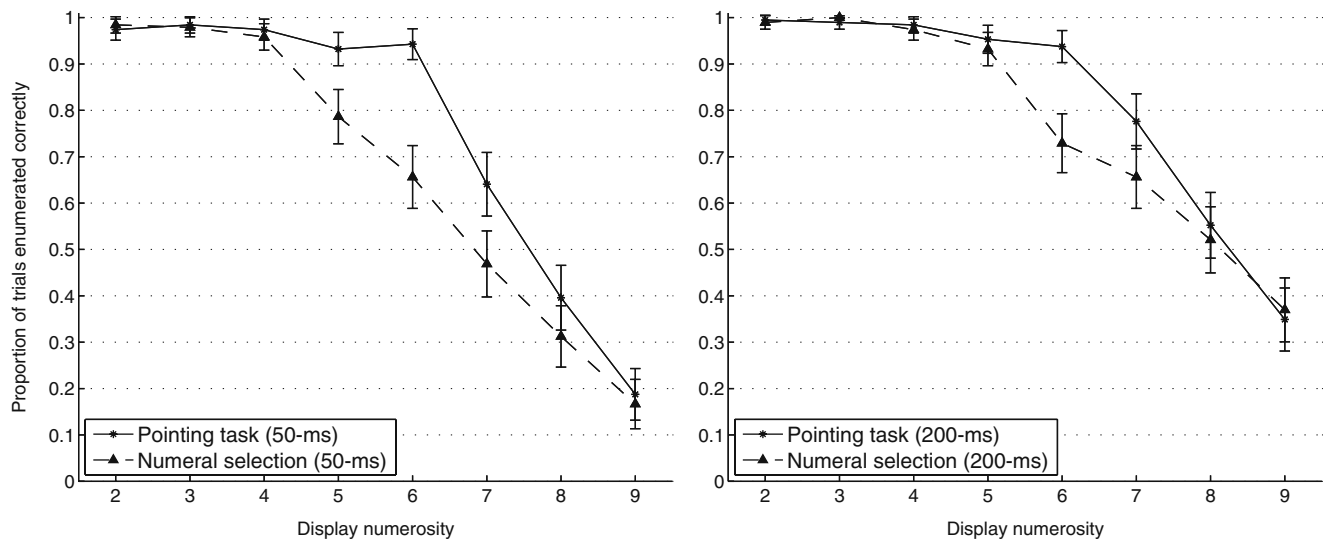


Fig. 5 Proportion of trials with perfect enumeration performance for pointing and numeral report conditions in Experiment 2 (50-ms on *left* and 200-ms on *right*)

(Figs. 4 and 6), whereas “enumeration” performance only decreased when there were more than six objects (Figs. 2, 3, and 5). These findings suggest that: (1) observers’ enumeration performance is based on items that they had individuated rather than on a strategy that uses some global property of the display (such as the total area of black discs); (2) observers are able to correctly individuate and report objects even when their report of locations was relatively impaired (i.e., the increasing location errors in the 2–6 item condition); and (3) with location responses such as those used here, observers can recall up to six individual objects.

There are several possible explanations for the larger subitizing range observed with the pointing response method. For example, the larger number of items recalled might be due to the use of motor “pointing” gestures. There is evidence that location information may be available for

accurately executing motor gestures even when it is not available for verbal report, and vice-versa (Goodale & Milner, 2004). Therefore, the pointing response used in our experiment may tap into a different system of (motor) representation, which in turn leads to external markings that could be used by the symbolic counting process.

Another possible account of the difference between these two methods relies on Visual Indexing Theory. This theory (Pylyshyn, 1989, 2001, 2007) proposes a limited set of indexes that automatically pick out individual visual objects. The indexing mechanism does not itself encode object properties nor does it provide a numerical code for the cardinality of the set of indexed items. It merely provides an indexical reference to the individual objects so that subsequent processes can operate on them. Thus, to derive the cardinality of the set of indexed objects, a subsequent stage of enumeration is required. When there are fewer objects than the indexing limit, enumeration operates over already-individuated and indexed items, rather than over the original display, so it bypasses the slowest aspect of counting (i.e., finding, individuating and marking objects) that must be used to enumerate larger sets. This account is consistent with the finding that the pattern of performance may be affected distinctly by different factors since some factors may affect the first (indexing) stage and others may affect the second (counting) stage. Thus it is consistent with the observation that the “knee” or inflection of the performance curve in Figs. 2 and 3 appear to be shifted towards higher performance as the stimulus duration increases. This also applies to the results in Experiment 2, where the performance differences between the two reporting methodologies were more pronounced in the 50-ms displays. In such short durations, earlier processes such as individuation may be impaired (as reported by Lorinstein & Haber, 1975).

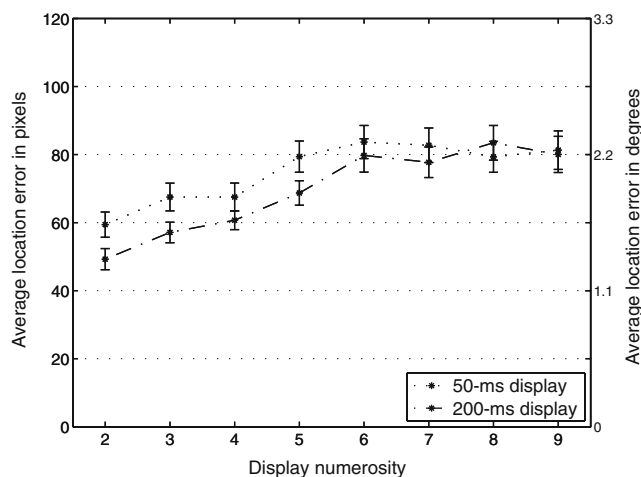


Fig. 6 Average distance between stimulus-response pairs (from Block 2 of Experiment 2), in pixels (*left y-axis*) and visual degrees (*right y-axis*)

If the subitizing limit is taken to be the largest number of individuals that are retained under ideal perceptual conditions, then we see from Fig. 2 that this number is approximately six items. However, the limit may also be taken to be the largest number of items that can be retained when viewing conditions are less favorable, for example when the display duration is short. Figure 2 shows that when the display is short, the resulting subitizing limit is smaller (around five). The problem with using this as an estimate of the subitizing limit is that the reduction in the robustness is due primarily to performance at the 50-ms stimulus duration. As noted earlier, performance at 50-ms may limit the earliest individuation stage which is only one part of the subitizing process.

Perhaps an even more promising account of why enumerating by pointing may be more efficient is that it may provide a way to keep track of items that have already been counted. In some cases, this may be done by clustering already-counted objects into mnemonic groups, which may be why grouping objects into canonical patterns improves the efficiency of enumeration (Mandler & Shebo, 1982). Another way to mark the already-counted items is available when the pointing method is used and if we assume that pointing benefits from the motor representation (via the dorsal visual stream). If the objects are no longer present, as in our experiment, using this motor representation to place marks on their former locations can help keep track of already-counted objects. As long as one can associate particular objects with particular marks (located with a precision at least as accurate as the inter-object spacing), the marks placed on the screen will provide a visible mark that can be used to identify already-counted objects.

Thus, there are several candidate hypotheses for how the pointing method might help to increase the span of recall or “enumerating” in subitizing experiments. These provide theoretical challenges as well as ideas for further experiments that may support one or another of these options.

Acknowledgements This research was supported by NSF IGERT #0549115 (H.H.H.) and Rutgers University institutional research funds (Z.W.P.). The authors thank Xiaotao Su (programmer), three anonymous reviewers, and Deborah Aks, Randy Gallistel, Manish Singh, and the Rutgers Visual Attention Lab for insightful discussions.

References

Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research*, 39(1), 71–86. doi:10.1016/S0042-6989(98)00090-X

- 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. *The Behavioral and Brain Sciences*, 24(1), 87–114.
- Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314. doi:10.1016/j.tics.2004.05.002
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44(1–2), 43–74. doi:10.1016/0010-0277(92)90050-R
- Goodale, M. A., & Milner, A. D. (2004). *Sight unseen: An exploration of conscious and unconscious vision*. Oxford: Oxford University Press.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, 43(3), 171–216. doi:10.1006/cogp.2001.0755
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *The 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. doi:10.1007/BFb0033292
- Klahr, D. (1973). Quantification processes. In W. G. Chase (Ed.), *Visual information processing* (pp. 3–34). New York: Academic Press.
- Lorin, B., & Haber, R. N. (1975). Perceived numerosity: An information processing analysis. *Canadian Journal of Psychology*, 29(3), 224–236. doi:10.1037/h0082028
- Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, 111(1), 1–22. doi:10.1037/0096-3445.111.1.1
- 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. doi:10.1037/0097-7403.9.3.320
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the first spatial-index model. *Cognition*, 32(1), 65–97. doi:10.1016/0010-0277(89)90014-0
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1–2), 127–158. doi:10.1016/S0010-0277(00)00156-6
- Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world*. Cambridge, MA: MIT Press.
- Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2), 331–351. doi:10.1037/0096-1523.19.2.331
- Trick, L. M., & Pylyshyn, Z. W. (1994a). Cueing and counting: Does the position of the attentional focus affect enumeration? *Visual Cognition*, 1(1), 67–100. doi:10.1080/13506289408402294
- Trick, L. M., & Pylyshyn, Z. W. (1994b). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, 101(1), 80–102. doi:10.1037/0033-295X.101.1.80
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10(2), 130–137. doi:10.1111/1467-9280.00120