ENUMERATING BY POINTING TO LOCATIONS:
THE ROLE OF SPATIAL INFORMATION IN COUNTING

By
HARRY HAROUTIOUN HALADJIAN

A Dissertation submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Psychology
written under the direction of
Dr. Zenon W. Pylyshyn
and approved by

________________________
________________________
________________________
________________________

New Brunswick, New Jersey

May, 2011
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. The underlying mechanism for small-set enumeration was examined by using a novel “pointing” task that also obtains information about 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. The observers then placed markers on a blank screen to indicate the locations of these discs. The number of these markers provided an estimate of enumeration accuracy. Observers could “enumerate” accurately displays containing up to six items in contrast with the four-item limit typically found. Experiment 2 replicated this lower limit by asking observers to report numerosity using cardinal symbols. Here, enumeration performance was better in the pointing task than the numeral report task. Experiments 3-5 presented alternative control tasks in addition to the pointing task. Results from these experiments indicate overall better enumeration performance in the pointing task, which suggests a non-symbolic response method produces a higher subitizing range.
To characterize the mechanism underlying the pointing task, we examined the errors in the memory for object locations. Localization error was measured as the distance between corresponding stimulus discs and response markers. These errors increased in magnitude with larger numerosities and shorter display durations. Responses also indicated a compression of distances around the center of mass. Additionally, stimulus displays with higher regularity in the spacing between discs produced better enumeration and localization performance. Overall, the localization results indicate spatial information can be represented with a rough accuracy that improves when increasing exposure time to the stimulus or reducing the number of items that must be remembered. In contrast, enumeration performance showed few errors in the same range (2-6 items). Together, these results support the possibility that the Visual Indexing mechanism assists small-set enumeration by individuating and selecting discrete objects without necessarily encoding detailed information about these objects.
Acknowledgements

This research was supported by NSF IGERT #0549115 (2008-2010) and Rutgers University institutional research funds (awarded to Dr. Zenon Pylyshyn at the Rutgers Center for Cognitive Science). I would like to thank my dissertation committee—Zenon, Randy, Manish, and Fabien—for sharing their wisdom with me. I also wish to thank Xiaotao Su for programming assistance, and the Object Group and Visual Attention Lab for insightful discussions about this work, with special thanks to Deborah Aks, Ryan McKendrick, Brian Keane, and Carlos Montemayor. Completion of this thesis would not have been possible without support by my friends and family, namely Chad Miller, my sister Caroline, and my mother Georgette. Other noteworthy sources of support for my doctoral thesis include Lupe’s, Kaffe 1668, and the members of SG-1.
# Table of Contents

Abstract .............................................................................................................................. ii

Acknowledgements ........................................................................................................... iv

List of Figures & Tables ................................................................................................... vi

Chapter 1. Introduction ...................................................................................................... 1

Chapter 2. Mechanisms of Enumeration ............................................................................ 9

Chapter 3. Study Goals .................................................................................................... 52

Chapter 4. Experimental Methods & Results ................................................................. 57

   Experiment 1 ........................................................................................................ 64

   Experiment 2A ..................................................................................................... 72

   Experiment 2B ..................................................................................................... 81

   Experiment 3A ..................................................................................................... 90

   Experiment 3B ..................................................................................................... 97

   Experiment 4 ...................................................................................................... 104

   Experiment 5 ...................................................................................................... 110

   Discussion for Experiments 1-5 ......................................................................... 118

   Analysis of the Localization Data ..................................................................... 121

Chapter 5. General Discussion ....................................................................................... 152

Chapter 6. Conclusions .................................................................................................. 165

Appendix ........................................................................................................................ 169

Bibliography .................................................................................................................. 177

Curriculum Vita ............................................................................................................. 189
List of Figures & Tables

Figure 1. Schematic of the enumerating-by-pointing task .............................................................. 64
Table 1. Regression values for Experiment 1 ...................................................................................... 69
Figure 2. Proportion of trials enumerated correctly in Experiment 1 .............................................. 70
Figure 3. Average miscounts in Experiment 1 .................................................................................. 70
Figure 4. Average distance between stimulus-response pairs in Experiment 1 ......................... 71
Figure 5. Depictions of the response screens used in Experiments 2A and 2B ......................... 74
Table 2. Regression values for Experiment 2A ............................................................................... 78
Figure 6. Proportion of trials enumerated correctly in Experiment 2A .......................................... 79
Figure 7. Average miscounts in Experiment 2A ............................................................................ 79
Figure 8. Response times in Experiment 2A .................................................................................. 79
Figure 9. Average distance between stimulus-response pairs in Experiment 2A ..................... 80
Table 3. Regression values for Experiment 2B ............................................................................... 87
Figure 10. Proportion of trials enumerated correctly in Experiment 2B ....................................... 87
Figure 11. Average miscounts in Experiment 2B ........................................................................ 88
Figure 12. Response times in Experiment 2B ............................................................................... 88
Figure 13. Average distance between stimulus-response pairs in Experiment 2B .................... 89
Table 4. Regression values for Experiment 3A .............................................................................. 94
Figure 14. Proportion of trials enumerated correctly in Experiment 3A ....................................... 95
Figure 15. Average miscounts in Experiment 3A ....................................................................... 95
Figure 16. Average distance between stimulus-response pairs in Experiment 3A .................... 96
Table 5. Regression values for Experiment 3B .............................................................................. 101
Figure 17. Proportion of trials enumerated correctly in Experiment 3B ..................................... 102
Figure 18. Average miscounts in Experiment 3B ....................................................................... 102
Figure 40. Average segment lengths of Delaunay triangle simplexes .............................................. 144
Figure 41. Mean localization error as a function of original display variability .............................. 146
Figure 42. Mean counting errors as a function of original display variability ...................................... 147
Figure 43. Coefficient of variance for the pointing task on 50-ms displays ........................................ 148
Figure 44. Coefficient of variance for the pointing task on 200-ms displays ........................................ 148
Figure 45. Coefficient of variance for the pointing task on 350-ms displays ........................................ 149
Appendix Figure A. Proportion of trials enumerated correctly in Exp. 2A, by participant .................. 169
Appendix Figure B. Proportion of trials enumerated correctly in Exp. 2B, by participant .................. 170
Appendix Figure C. Proportion of trials enumerated correctly in Exp. 3A, by participant .................. 171
Appendix Figure D. Proportion of trials enumerated correctly in Exp. 3B, by participant .................. 172
Appendix Figure E. Proportion of trials enumerated correctly in Exp. 4, by participant .................... 173
Appendix Figure F. Proportion of trials enumerated correctly in Exp. 5, by participant .................... 174
Appendix Figure G. Average miscounts in Experiments 1-5, by participant ................................. 175
Appendix Figure H. The cumulative distribution for enumeration accuracy ................................. 176
Chapter 1. Introduction

Attending to distinct visual objects is a crucial function of human cognition. The ability to identify and locate a number of objects serves numerous activities, such as retrieving target items, navigating through crowded environments, or playing team sports. Counting also depends on this ability, where accurate enumeration requires the selection of discrete perceptual objects and the assignment of a cardinal label to this set. Related to this sort of numerical perception are the processes of estimating quantity, representing duration, and extracting statistical information from visual scenes. Enumeration abilities have been examined extensively through studies on animals, infants, and adults, but whether a single mechanism or multiple systems are responsible for enumeration remains debated. The aim of the current project is to address this issue by focusing on the selection component of enumeration, that is, by examining how processes in early vision individuate and select visual items for enumeration.

A primary system that detects the number of items in a set—its numerosity—is the magnitude estimation mechanism. This mechanism also supports the internal representations of duration (Meck & Church, 1983), and typically produces more errors as the reference numerosity increases. Number “labels” may be mapped to discrete units on this continuous representation, and thus may account for verbal counting in humans (Gallistel & Gelman, 1992). When observers enumerate smaller sets of items, they make few errors with minor increases in reaction times (RT), but this rate of errors and RT increases substantially for sets larger than four (Trick & Pylyshyn, 1993, 1994b). The term *subitizing* refers to the quick and accurate enumeration of small sets (Kaufman, Lord, Reese, & Volkmann, 1949). The distinct performance for small-set enumeration
has led some to argue that there is a separate system—in addition to the magnitude estimation mechanism—that can derive the numerosity from a set of visual items.

Subitizing may be a fast counting process achieved via a fast mapping of cardinal labels to a magnitude representation (Gallistel & Gelman, 1992). This process is easier for smaller numerosities since the items are easier to discriminate and select for enumeration. Subitizing also may be the result of a fast counting of active visual indexes, which are part of a mechanism in early vision that establishes referents to four or five visual objects (Pylyshyn, 1989; Trick & Pylyshyn, 1993). The indexing mechanism is thought to be responsible for the individuation and selection of visual objects, and thus, will be active when selecting sets with up to four or five objects. Both of these proposals require a serial component in order to determine numerosity, for example, by a serial counting or incrementing process.

Another perspective that accounts for the fast enumeration of small sets argues that this process is a direct apprehension of numerosity. This visual “sense of number” may be served by a specialized mechanism, one that is parallel, preattentive, and unconstrained by perceptual capacity limitations (Sagi & Julesz, 1984). A direct apprehension of numerosity may occur via an automatic mapping between the magnitude estimation mechanism and simultaneous neural activations produced from a parallel detection of up to five objects (Dehaene & Changeux, 1993). Others propose that subitizing is a pattern-recognition process, where familiar patterns formed by a small number of items are recognized more quickly, for example, like the patterns on dice (Mandler & Shebo, 1982). The shared assertion in these proposals is that there is a direct association between sets stored in memory and the label for that numerosity. Therefore,
the numerosity can be detected instantly without the need to count serially. Such accounts for subitizing will be discussed in more detail in Section 2.

Regardless of the exact mechanism that enables enumeration, attention must be involved in some way to produce a numerosity judgment. For example, counting large sets of items requires attending to each to-be-counted object and avoiding already-counted objects—a resource intensive operation within focal attention. Even fast processes, such as estimation and subitizing, may be sensitive to attentional load (Egeth, Leonard, & Palomares, 2008; Olivers & Watson, 2008). A task-dependent engagement of the various forms of attention, such as spatially distributed or narrowly focused attention, may determine which mechanism is used during enumeration, or vice versa. For example, rapid preattention automatically occurs in early visual processing prior to focused attention and accounts for rapid selection and individuation abilities, which may include subitizing (Trick & Pylyshyn, 1993, 1994b). Top-down processes modulate a more resource-intensive focused attention in order to perform difficult cognitive tasks, such as detecting scene changes (Rensink, 2000) or counting serially (Trick, 2005). Attention, therefore, can be characterized as having both automatic bottom-up and controlled top-down operations that enable an agent to process information from the environment in a meaningful manner (Cave & Bichot, 1999; Connor, Egeth, & Yantis, 2004; Treisman, 2006; Watson & Kramer, 1999).

There are at least two functions involved in attention that are crucial for enumeration: individuation and identification. The first of these, individuation, is a data-driven process that selects objects for further processing and is thought to occur within a cognitively impenetrable module referred to as “early vision” (Pylyshyn, 1999).
According to Ullman (1984), “visual routines” may assist individuation and other processes by extracting spatial relations and other properties from the output of early visual processing. Such visual routines assist in tracking contours, counting objects, or marking locations. This sort of bottom-up processing must operate on a scene before features can be encoded and is important for understanding the mechanism that may support subitizing (Trick, 1992). The second stage “binds” the features of attended objects into richer mental representations and allows object identification or recognition (Ballard, Hayhoe, Pook, & Rao, 1997; Kahneman, Treisman, & Gibbs, 1992). Under the rapid counting proposal of subitizing, it is during this stage where a numerical label can be assigned. Together, individuation and identification can produce the experience of attending to specific items in the world, including when enumerating sets of items.

The first stage of object individuation relies on the coherence of visual input that extends over space and time (Feldman & Tremoulet, 2006; Scholl, Pylyshyn, & Feldman, 2001). This is influenced by the spatial relationships among visual stimuli, including geometric factors such as symmetry, good continuation, and parallelism (Feldman, 2007). Segmentation processes in early vision likely operate to divide visual arrays into distinct objects in accord with the principles of cohesion, boundedness, rigidity, and “no action at a distance”, or that distant objects appear to be independent from one another (Spelke, 1990). By dividing visual scenes into distinct objects, an organizing structure is formed that enables binding features in visual short-term memory, which is dependent on attention (Feldman, 2003; Wheeler & Treisman, 2002). This individuation process plays a crucial role in subitizing.
The mechanism responsible for individuating visual objects is described by Visual Indexing Theory (Pylyshyn, 1989, 2001). Visual indexes are “pointers” to contextual items in a visual scene (also named “FINSTs” for “fingers of instantiation”). These data-driven and preattentive pointers are triggered by certain object properties, such as cohesion or rigidity. Once a pointer is assigned, it can “stick” to and follow a visual object (Pylyshyn, 1989, 2003, 2007). A paradigm for testing the visual indexing mechanism is the multiple object tracking (MOT) task, where observers track several target objects that move unpredictably and independently among identical distractors (Pylyshyn, 1988, 2001). Up to four or five of these moving objects can be tracked successfully for extended durations under varying conditions, including when targets move behind occluders or change in shape and size.

Attention can be shifted among visual indexes depending upon task demands (Pylyshyn, 1989), and this attention can help build richer representations by integrating object features (Treisman, 1998; Treisman & Gelade, 1980). Object File Theory is a complementary theoretical framework to Visual Indexing Theory that provides an account of how detailed object representations are formed through the binding of featural information (Ballard, et al., 1997; Kahneman, et al., 1992). Selective attention plays a crucial role in forming persisting object representations by allowing features from a visual scene to build a coherent representation incrementally (Treisman, 1998, 2006). This binding happens quickly, with 200 ms being sufficient for encoding substantial information (Feldman, 2007). This construction occurs in visual short-term memory, which is limited in the number of objects and the amount of featural information that can be represented (Alvarez & Cavanagh, 2004; Xu & Chun, 2006). Object File Theory was
developed from experiments that showed a preattentive object-specific preview benefit, where observers responded faster to a target stimulus when it reappeared in the same object that briefly previewed it, even after moving through space (Kahneman, et al., 1992). This result emphasized the object-specific nature of the preview effect (as opposed to being location-specific) and became the basis for positing object file representations. These object files have also been used to explain the ability to count or subitize sets of objects (e.g., Carey, 2001; Feigenson & Carey, 2003).

The object-based attention model claims that the visual system favors the selection of whole objects in a visual scene as opposed to properties such as location or any other features, and many studies provide substantial support for an object-based theory of attention (Baylis & Driver, 1993; Kahneman, et al., 1992; O'Craven, Downing, & Kanwisher, 1999; Scholl, 2001; Treisman & Zhang, 2006; Yantis, 1992). Without the ability to select an individual object and bind its features, an agent could not sustain a persisting representation of that object (since features are features of objects and not of locations). Prior to the encoding of features, however, a discrete object must be parsed from a visual scene and identified with a place-holder or pointer (Pylyshyn, 1989). In fact, the individuation of visual objects is thought to be the first step in solving the binding problem, which is the problem of how several features can be integrated to form a coherent object representation (Treisman, 1996; for a review, see Wolfe & Cave, 1999). Several infant studies provide support for this type of feature integration into meaningful representations, but only after a certain developmental stage (Xu & Carey, 1996). Additionally, there may be an important disassociation between individuation and identification. For example, a study on 12-month-old infants found that feature
information used to individuate objects is not always included in its representation (Tremoulet, Leslie, & Hall, 2000). The finding that features can help individuate objects but do not need to be encoded into a persisting representation of this object is consistent with Visual Indexing Theory: the indexing mechanism operates prior to the deployment of selective attention and therefore can be active prior to the encoding of features related to an indexed item.

It is the early processing stage of individuation that is especially important for identifying how distinct visual objects can be selected for the purpose of enumeration (Trick & Pylyshyn, 1994b). Pylyshyn (2007) argues that once a small number of objects are individuated and indexed, their cardinality can be determined by counting the number of active indexes and, therefore, can be accomplished quickly since there is no further need to find and mark the items in the stimulus for counting or somehow segregate them in order to prevent recounting. Presumably, this process does not require scanning the display and individuating objects but simply counting the active indexes (independent of the features encoded).

To further characterize the mechanism for object individuation and its role in enumeration, the current study examines the process by which the cardinality of a set of simple visual items is attained. A number of different factors (such as timing and location of objects) are examined to determine how spatiotemporal factors affect enumeration. Additionally, the effect of the manner in which the observers reported numerosity is examined. A novel way of measuring observers’ implicit knowledge of numerosity was developed for this study consisting in recording the number of object locations that were reported by the observers. In this “pointing task”, observers marked the location of each
object that was briefly presented on a masked stimulus screen. This measure of “implicit” numerosity relies on the individuation stage of enumeration and provides a characterization of the type of spatial information involved in this process. The results from this method should indicate if the ability to report the numerosity of small sets can be characterized by an individuation mechanism in early vision, one that is preattentive, parallel, but limited in capacity, or if further processing is required in order to represent accurate numerosity that includes information about objects locations.
Chapter 2. Mechanisms of Enumeration

Several proposals have been made to account for subitizing. Some argue that subitizing relies on a primitive form of numerosity perception: continuous magnitude representations. Results from some studies, however, suggest that another system may process small numerosities, such as the visual indexing mechanism or a pattern-matching process. The vast literature on numerical cognition provides support for both single-mechanism and multiple-mechanism accounts of subitizing.

2.1. Types of Quantification Processes

Several terms are used to describe the process by which people arrive at the cardinality of a set of items. The fundamental concept in this process is number, which is an inherent property of sets of individuals that exists prior to any sort of numerical cognition (Dehaene, 1992; Kitcher, 1975). Number is an abstract concept of numerosity and does not need to refer to a specific set of items; it is a property shared by an unlimited number of sets that can be mapped onto one another by a bijective function. In this way, numbers can serve as the basis for conceptual (i.e., symbolic) systems, such as mathematics. The term numerosity has been used to indicate “a measurable numerical quantity” that is an objective property of a denumerable set of physical entities, and hence refers to actual things in the world (Dehaene, 1992; Gelman & Gallistel, 1978). This quantity can be represented mentally as a numeron, which is a subjective mental symbol for a set’s numerosity (Gelman & Gallistel, 1978, p. 77). A general term for determining the number of items in a set and encoding it in terms of its numeron is quantification. Three primary quantification processes with distinct characteristics have
been identified: *estimation, counting* (nonverbal and verbal), and *subitizing* (Dehaene, 1992; Kaufman, et al., 1949; Klahr, 1973).

### 2.1.1. Estimation

Estimation is an approximate derivation of numerosity based on a nonverbal analog magnitude representation. These representations are inferred mental entities that represent the numerosity or magnitude of items in the world and have the properties of real numbers (Gallistel & Gelman, 2005). A mental magnitude must have a causal connection to the quantity it represents, but this representation is independent of language abilities. Animal numerical representations, including duration, rely on such magnitude estimates (Gibbon, 1977). Meck & Church (1983) proposed an accumulator mechanism to account for the representations of duration and numerosity by rats. This study showed that a single “count” can be mapped on to a duration representation of 200 ms, which indicates that a fixed increment of a continuous number representation can correspond to a discrete unit. In humans, this system represents numerosities by employing a discrete incrementing process that defines the “next magnitude” by mapping a label to each successive discrete increment in a linear manner (presumably beginning with a standard smallest value). Such a process could then produce a verbal count (Gallistel & Gelman, 1992, 2000; Gelman & Butterworth, 2005). Arithmetic operations can be performed on magnitude representations, which explains a preverbal infant’s ability to add or subtract quantities (Gallistel & Gelman, 1992; Zur & Gelman, 2004) and an adult’s ability to use magnitude representations to compare numerosities without using verbal labels (Balci & Gallistel, 2006). From an evolutionary perspective, the presence of an innate accumulator explains how preverbal infants discriminate numerosities and provides a system that
allows the mapping of symbolic numerosity representations to magnitude representations (Uller, 2008).

Gallistel & Gelman (1992) propose a linear mapping of numerical values to magnitudes. The errors in these representations are proportional to the numerosity: over repeated trials, larger numerosities produce greater variability in the corresponding representations. The resulting coefficient of variation is characterized as a constant ratio between the standard deviation and the mean magnitude (Gallistel & Gelman, 2000; Gibbon, 1977). This scalar variability suggests that the discriminability between two magnitudes depends on their ratio in accordance with Weber’s Law: the larger the Weber fraction, the more difficult it is to discriminate between two numerosities (Gallistel & Gelman, 1992; Whalen, Gallistel, & Gelman, 1999). Similarly, Dehaene and colleagues propose a mental number line with an analog encoding of numerosities (Dehaene, 1989; Dehaene, Dupoux, & Mehler, 1990). A distinct feature of this model is that it explains the Weber-like characteristics of enumeration through a logarithmic mapping from the objective numerosity to positions on the number line instead of equal intervals on a linear scale. Both of these models provide accounts of how magnitudes can be represented (but an adequate discussion of the advantages and disadvantages of these two models is beyond the scope of the current project). This estimation mechanism also is thought to support verbal counting, but it remains unclear how the mapping between a magnitude and a numeron occurs.

2.1.2. Verbal counting

Verbal counting is a common way to enumerate perceptually distinct items, such as a set of objects or sounds. When counting visual items, they typically must be
individuated so that a discrete item is associated with a particular object, and already-
counted items must be “tagged” to prevent double-counting. The ability to count verbally
is thought to develop from associating a numerosity with a discrete increment on the
(continuous) mental number line and from the bidirectional mapping between this
increment and a numeron that represents the numerosity (Gallistel & Gelman, 1992). The
count-list numerons have a stable order and represent cardinality, where the last numeron
corresponds to the set’s numerosity. Counting can be abstracted to any set of entities and
the order of items to be counted in the set does not matter as long as errors in cardinality
are avoided (Gelman & Gallistel, 1978, p. 80). The crucial point is that cardinal-counting
requires mapping a one-to-one correspondence between items and their symbolic label,
and therefore, relies on the individuation process in early vision.

Verbal counting typically involves a serial selection of each item in a set and may
include reciting cardinal labels (e.g., “one”, “two”, “three”). This raises the question of
whether counting always relies on language. The language faculty allowed humans to
develop number notations and symbols used in complex mathematical computations
(e.g., calculus), but the ability to quantify or approximate numerosities does not rely on
manipulating linguistic labels and is present in animals and preverbal infants (Dehaene,
1992). A view of verbal counting holds that verbal representations of numerosity map on
to preverbal mental magnitudes, and learning this mapping corresponds to learning how
to count (Gallistel & Gelman, 1992). A symbolic mapping to magnitudes was shown by
Moyer & Landauer (1967) in experiments that asked observers to judge the larger of two
digits. Two robust results were obtained: larger numerical differences between two digits
produced faster and more accurate responses (“distance effect”), and large numerosities
produced slower responses with more errors ("size effect"). Since these results resembled those obtained when humans discriminate between physical quantities, such as the length of a line or tonal pitch, the authors concluded that numerosity judgments first required a conversion of numerical symbols into analog magnitudes prior to comparing them.

There are alternative proposals, however, that describe the learning of count words as being independent of magnitude representations. One such view attributes the learning of number words to the representations produced by a parallel individuation system that stores information about individuals (Carey, 2004; Le Corre & Carey, 2007). Numerosity can be derived from these representations of individuals and can be stored in working memory (and thus subject to working memory capacity limitations). Studies continue to address the emergence of verbal counting in humans and how this relates to a more "primitive" nonverbal enumeration system. (The effect of language on enumeration is discussed further in Section 2.2.5.)

2.1.3. Subitizing

Subitizing, as previously mentioned, describes a quick and accurate enumeration of small sets containing around four items. Beyond the subitizing range of four items, enumeration accuracy drops and RT increases more rapidly with increasing numbers of items. Results from various studies indicate that the RT for enumerating small sets typically increases at a slope of 10-100 ms per item, and the slope increases substantially to about 250-350 ms in larger sets (as reviewed in Trick & Pylyshyn, 1994b). The change in the RT slope also may be an indication of a shift in strategy (e.g., from estimation to serial counting) or of a separate cognitive process, such as the visual indexing mechanism. This mechanism typically can individuate and select up to four items in
parallel, and numerosity can be determined by some sort of rapid serial counting process on these active indexes (Folk, Egeth, & Kwak, 1988; Trick & Pylyshyn, 1994b). Alternatively, the strategy may shift to a pattern-matching process (Mandler & Shebo, 1982); this view attributes the small RT increases within the subitizing range to the increasing number of familiar patterns to search through (in memory) in order to produce a corresponding numerosity match. For sets with more than three or four items, the strategy may also shift to a slower mental counting process that increases the demands for attention (e.g., by requiring to keep track of what has and has not been counted).

Whether there is a true performance discontinuity when enumerating small and large sets, however, continues to be debated. The view that subitizing is simply due to a fast counting process (that does not rely on a parallel selection of the items to be counted) asserts that for every additional object that must be enumerated, there is an increase in RT that is substantially greater than the preceding increment, which is a result of counting (Gallistel & Gelman, 1991). On this view, errors in enumeration grow in proportion to the numerosity. The subitizing effect also may be due to the ease of individuating scenes with few objects or the fast mapping of symbols to small magnitudes. The various studies on and explanations for subitizing will be examined further in the following sections.

2.2. Is There a Distinct Mechanism for Subitizing?

Does the larger increase in the rate of errors and RTs with each added item when enumerating larger sets indicate separate mechanisms for enumeration or different demands upon a single magnitude representation system? Despite the abundance of studies addressing this question, it remains debated whether one or multiple systems serve enumeration.
A single mechanism account of enumeration is primarily supported by the presence of a constant coefficient of variation during enumeration: the variability in responses increases as the cardinality of the set to be quantified increases. This scalar variability is observed in animals and humans, and suggests a shared mechanism (Gallistel & Gelman, 1992, 2000). One study tested this shared mechanism account by using two methodologies typically administered to animals: the key-press and flash-count paradigms (Whalen, et al., 1999). In the key-press condition, observers viewed Arabic numerals (odd numbers from 7-25) and pressed a computer key the corresponding number of times as fast as possible. In the flash-count condition, observers viewed a number of dots in rapid succession (controlled for duration) and estimated the numerosity of dots in each trial. Both experiments yielded results similar to the previously reported animal data: the numerosity estimates of larger sets produced increasing standard deviations, resulting in a constant coefficient of variation. To ensure that verbal counting was not used in this task, the observers’ RTs were confirmed to be faster than the required duration for sub-vocal counting. In addition, response patterns did not indicate that chunking strategies were used to make responses (i.e., RTs did not suggest grouping of key presses). These results were extended by another study on verbal and nonverbal counting where observers quickly pressed a button a number of times to match a numeral probe (Cordes, Gelman, Gallistel, & Whalen, 2001). In the nonverbal counting condition, observers said “the” with every key press to prevent verbal counting. In the verbal counting condition, they counted audibly with every press either serially up to the numerosity or in sets of ten to control for the number of syllables (i.e., after 10 the count restarted at 1). Scalar variability was observed only in the nonverbal condition,
suggesting that a mental magnitude generates representations of small and large numerosities during nonverbal enumeration.

Another explanation for fast enumeration abilities is that observers make judgments on a feature of the display, such as the amount of space the objects occupy, rather than enumerating individual items. To test this possibility, a recent study presented observers with a display containing two arrays of discs with varying numerosities and total surface areas (Hurewitz, Gelman, & Schnitzer, 2006). In two experiments, observers were asked to make speeded judgments identifying the array with more discs or to identify the array with discs occupying the greatest combined area. Disc size and number were manipulated in order to control for the total area covered by the discs on the displays to test whether continuous information about surface area facilitated enumeration or if it was driven by discrete items on the display. The results showed that enumeration was influenced by the relative size of the discs regardless of set size (however, it should be noted that Trick & Pylyshyn, 1994a, did not find an effect of variable disc sizes on enumeration accuracy). Additionally, continuous judgments (i.e., when indicating which array had discs that covered more space) were faster than discrete judgments (i.e., indicating which array had more items). Such results describe an hierarchical structure in numerical processing, where continuous numerosity judgments are less cognitively demanding (and more automatic) than serial counting. A greater sensitivity to continuous quantities is also supported by findings from an infant study that indicate continuous representations are available prior to discrete number concepts (Xu & Arriaga, 2007). Other infant studies suggest, however, that both discrete numerosity and the magnitude of the cumulative surface area can be represented without an overriding
preference for one form of representation (Cordes & Brannon, 2008b, 2009). In any case, there appears to be a developmental relationship between these abilities that implies a common nonverbal estimation mechanism present innately for numerical cognition.

The argument for a single enumeration mechanism relies on the presence of errors that follow Weber’s Law (for all numerosities). Operating under this assumption, a recent study tested whether or not a subitizing task produced the same scalar variability as large numerosity estimates (Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). Observers judged numerosity by identifying which of two displays had more dots in one experiment and by naming the numerosity in another experiment. The displays appeared briefly (150 ms) and had small sets (1-8) or large sets (10-80) of black dots on a white background. The larger sets were presented in “decade quantities” (e.g., “10”, “20”, “30”), on which observers were trained. The training involved a calibration process that improved enumeration performance simply by telling observers the number of dots that a display contained, and the observers could calibrate their responses to be more accurate in subsequent trials (described in Izard & Dehaene, 2008). The area occupied by the dots was held constant in half of the trials to prevent the usage of non-numerosity cues. The authors reasoned that if the same mechanism enumerates all numerosities, then the scalar variability across the ranges in the small-set and large-set display conditions should match since the ratio of the numerosities was equivalent between these two displays. The resulting errors and RTs showed that performance in the subitizing range was more accurate and faster than in the counting range, so the scalar variance ratio was different in the two ranges. This violated Weber’s Law predictions of discriminability since a distinct scalar ratio was observed only for the range of 1-4 (out of 8) items and not in displays
with 10-40 (out of 80) items. The authors argue that these results support a distinct system for perceiving small numerosities.

There are strong arguments for multiple “core systems” of numerical representations that can be observed in animals and tracked developmentally in humans. Spelke & Kinzler (2007) propose five possible core knowledge systems, two of which are responsible for number representations—one for approximate magnitude estimates and another for representing discrete individuals. Enumerating small sets may engage a preattentive, parallel individuation mechanism, and the number of these active indexes may be counted when enumerating smaller sets. These indexes also enable higher-level representations such as object files (as proposed by Kahneman, et al., 1992). Object files themselves do not have numerical properties like mental magnitudes, although representations of magnitudes may be stored in these files. Carey (2001) argues, however, that the operation of opening a new object file is analogous to adding the value of “1” to a representation. This process may help infants initiate numerical reasoning and could explain subitizing. This view may be problematic since sets of objects do not have numerical content without another representational system “acting” upon this set in order to detect its numerosity, for example, by counting the items. Furthermore, one would still need to specify how such a numerical representation is formed (Gallistel, 2007).

Trick & Pylyshyn (1989, 1993, 1994b) propose an account for subitizing by citing the characteristics of the visual indexing mechanism. Visual indexes are pointers that automatically detect and “stick” to visual items that have the qualities of “objecthood” (e.g., good continuation, cohesion, common fate). According to the Visual Index theory, indexing is a preattentive, automatic, and numerically limited parallel process. It has been
used to explain the ability in human observers to track multiple moving target objects among distractors, select subsets of items, and establish a causal link between items in the world and their mental representations (Pylyshyn, 1989, 2007; Trick & Pylyshyn, 1989, 1993, 1994a). Both Indexing Theory and the related Object File Theory postulate a capacity for maintaining around four indexes or object files simultaneously. These mechanisms, however, do not in themselves possess a representation of the cardinality of a set. For that, a numeron must be assigned by counting the active indexes. Subitizing is assumed to be a process by which a numeron is associated with the number of active indexes (or object files). This process is rapid because presumably active indexes are easier to count since these items have already been individuated. The (putative) discontinuity in performance between small and large-set enumeration is taken as support for the presence of two systems for enumeration (Feigenson, Dehaene, & Spelke, 2004).

Since by hypothesis subitizing does not require focused attention or feature binding (Trick & Pylyshyn, 1994a), Visual Indexing Theory provides a viable mechanism that explains the fast enumeration of 1-4 objects without the need for encoding specific object properties. This subitizing account assigns a numeron to the number of active indexes or object files, regardless of their representational content. In this sense, the activation of two indexes or object files is simply the activation of two tokens, and this *activation* is used to determine numerosity. The indexing mechanism has a capacity of four items and detecting the numerosity within this range can be fast, accurate, and effortless. Although indexes may be assigned in parallel, there appears to be an effect of serial counting since there is a minimal increase in RT when enumerating 1-4
items (roughly 10-100 ms per item). One must specify, however, the process by which the set of pointers is assigned a numerical symbol, and this remains an open question.

If subitizing is based upon a capacity-limited indexing mechanism, it could explain why four items can be enumerated with minimal attention. Larger sets require more attentional resources and alternative strategies for enumeration, such as counting via serial attention or an approximate estimation process. The preattentive nature of the indexing process has been supported by the finding that the locus of attention does not affect subitizing. Small sets are enumerated accurately even when the targets lie outside focal attention, but performance declines when five or more items are presented (Trick & Pylyshyn, 1994a). Easily subitized displays tend to involve pop-out features that reduce the need for serial attention to isolate these items (Feigenson, 2008; Trick & Pylyshyn, 1989, 1994b). Like pattern detection or MOT, this process is sensitive to crowding (Franconeri, Alvarez, & Enns, 2007). To facilitate pop-out, items must be sufficiently distinct from the background and occupy discrete and identifiable locations in space with minimal crowding. Since visual processes, such as apparent motion or stereo fusion, appear to operate on discrete units (Ullman, 1979), the individuation of these objects may be used for numerical processing (Brase, Cosmides, & Tooby, 1998; Franconeri, Bemis, & Alvarez, 2009). A recent study further supports the possibility that the visual indexing mechanism is responsible for subitizing by the finding that subitizing is especially responsive to moving stimuli (Alston & Humphreys, 2004). When enumerating displays with moving or static targets, observers showed better performance when the to-be-counted items were moving and especially poor performance when the displays had static targets and moving distractors.
The multiple-system perspective argues that although a magnitude representation is used in a variety of enumeration tasks, another system that quickly individuates a limited number of items also can serve the purpose of identifying numerosity since the individuals can be counted more quickly. The following sections highlight some of the findings from diverse studies that have implications on whether or not a separate individuation mechanism supports subitizing.

2.2.1. Display features and patterns

In addition to the pop-out quality of easily enumerated displays, other scene features can influence enumeration. Salient features can facilitate the individuation (and thus enumeration) of objects, but individual features cannot be enumerated rapidly in the same way that objects occupying distinct spatial regions are enumerated. Enumerating-by-feature, for example, when counting the number of colors on a display, is a slow and serial process that is distinct from the fast enumeration of individual targets (Watson, Maylor, & Bruce, 2005b). Therefore, object features are often ignored during enumeration, unless these features encourage grouping. For example, studies on young children have shown that they can maintain a distinction between numerosity and other display features (Gelman & Tucker, 1975), but they also can use spatial information to group arrays into sets for numerical processing (Feigenson & Halberda, 2004). Proximity is an important property for determining which items are more likely to be grouped together (Kubovy & van den Berg, 2008; Quinn, Bhatt, & Hayden, 2008; van Oeffelen & Vos, 1983). A study by van Oeffelen & Vos (1982) showed that large arrays (13-23) of dots were likely to be subitized in groups of four or less based on their proximity, and the partial results of each subitized group were added to a running total. Another study
confirmed that grouping can be used to enumerate quickly and can be a learned expertise, as observed in air traffic controllers (Allen & McGeorge, 2008). Grouping, however, does not always occur. The study by Whalen et al. (1999) did not find evidence for grouping in a nonverbal counting task that used key presses to obtain information about the observers’ numerosity representations. Grouping appears to be a task-dependent strategy that may rely on the availability of attentional resources.

A study on perceptual grouping by Trick & Enns (1997) used two types of stimuli in an enumeration task: one display with drawn squares and diamonds, and another with clusters of four dots that formed the squares or diamonds. In a simple enumeration task, connectedness did not affect the RTs for enumeration and observers performed identically on displays with drawn objects and clustered objects. When distractors were introduced (i.e., by enumerating only squares), enumeration suffered in the clustered displays since the square targets were kept distinct from the diamond distractors. The clustering of dots into “countable objects” was sufficient for enumeration, but to discriminate whether the dot clusters formed a diamond or a square required longer processing and this process impaired enumeration. This suggests that some aspects of grouping, like identifying shapes made from groups of dots, require attention and others can be more automatic, like proximity grouping. This also may be related to being able to attend to different spatial frequencies, which would affect how such clusters are perceived (He, Cavanagh, & Intriligator, 1997).

A series of experiments by Trick (2008) further explored the role of automatic grouping in enumeration. Observers were implicitly encouraged to adopt a “group and add” strategy when enumerating displays of 1-9 items that were comprised of 1-3 shapes
(red squares, blue triangles, and green stars). Specifically, they were instructed simply to enumerate the entire display but not told explicitly to enumerate similar subsets and add their numerosities together. These results were compared to displays with the same numerosity but containing only one type of shape. The displays with heterogeneous shapes, which encouraged the grouping and enumeration of smaller sets, were enumerated significantly faster than the equivalent homogenous displays. A sharp change in the RT slope between 1-3 items and 4-8 items was evident in all of these experiments, including displays where items could change shape or color during the trial. This performance was explained by the indexing mechanism that rapidly selects a small set of objects in parallel (by assigning “FINSTs”), and the numerosity (i.e., number label) of these active indexes can be transferred into working memory in order to compute the overall numerosity. Grouping based on similarity allows for the visual system to exploit the parallel individuation mechanism and visual scene statistics, but the bottleneck of information transfer still occurs during the adding stage, which is susceptible to the capacity limitations inherent to working memory.

Grouping can interact with pattern matching to facilitate enumeration. When items are grouped in a way to produce familiar figures, such as squares or triangles, the subitizing range increases from the expected limit of four items. This benefit was also observed with familiar dot patterns like those found on dice. Mandler & Shebo (1982) found that grouping by triplets occurred when enumerating nine items in a canonical display, with errors significantly lower than eight- or ten-item displays (using 200-ms masked presentations of dot stimuli). These results support a strategy of grouping into three-item sets and adding these “chunks” to obtain numerosity. Grouping likely occurs
in this situation since it has been shown to occur in under 120 ms, where 87 ms allows proximity grouping and 118 ms allows alignment grouping (Kurylo, 1997). Mandler & Shebo (1982) proposed that subitizing is the recognition of common or canonical patterns, such as doublets and triplets. To test the relationship between subitizing and sensory memory for pattern recognition, observers were presented with different types of dot patterns (e.g., canonical, grouped, random) at different durations. In analyzing the types of errors made when enumerating 1-7 items, a robust asymmetrical error pattern emerged where sets of 1-4 were overestimated while larger sets were underestimated. This contrasts their hypothesis that using a single enumeration mechanism would produce symmetrically-distributed errors. The individuation of items in the subitizing range may be a “pop-out” of a memorized pattern, but larger sets require a serial process that increases RT and is prone to more noise. The reason that a pattern-matching advantage is not observed for larger sets is because the possible number of patterns increases exponentially and makes it intractable to determine numerosity this way.\(^1\)

Other support for the pattern-matching hypothesis is the observation that the similarity between dot patterns plays a role in identifying numerosities. When observers are trained on patterns of dot displays, larger sets can be recognized faster and this learning can transfer to other similar patterns (Palmeri, 1997). This pattern-matching account of subitizing posits that the target displays act as retrieval cues that activate numerical responses associated with similar displays. The slight increase in RTs within

\(^1\) Ullman (1984) argues that many patterns (e.g., ones requiring detecting the “inside” relation, or of detecting whether a contour is continuous) cannot be recognized without the serial application of visual operations in processes called “visual routines”. Moreover, the individuation of relevant primitives, as well as mapping a display into a canonical scale and orientation, must occur before any “lookup” can take place. So what may seem like a simple operation (e.g., noting that the items form a rectangle) is in fact quite complex.
the subitizing range may be explained by easily accessing the stored patterns (i.e., with more dots, there are more possible patterns that can be stored in memory). A recent study by Logan & Zbrodoff (2003) asked observers to rate the similarity between two displays with equal or unequal numbers (1-10) of white asterisks, which were randomly placed within an 8x8 grid on a black background. When shown displays with 1-3 items, observers correctly reported higher similarity in same-numerosity displays and lower similarity in different-numerosity displays, but accuracy decreased for displays with more than three items. The authors implemented a “generalized context model” (Nosofsky, 1984) to test whether their results indicate a direct-access memory retrieval and found a higher probability for memory retrieval when processing displays with sets in the subitizing range. This suggests that the primary characteristics of subitizing (i.e., being fast and automatic) follows from theories of automaticity and direct-access memory retrieval, which could indicate that subitizing is a similarity judgment that relies on retrieval from long-term memory (Logan & Zbrodoff, 2003). These results should be examined critically, however, since rating similarity is not necessarily the same task as reporting precise numerosity. Additionally, an alternative interpretation of such results would favor the sensitivity to scene regularities, such as spacing, and not necessarily a pattern-matching process per se (e.g., see Brady, Konkle, & Alvarez, 2009).

If enumerating small sets is indeed a pattern-matching task rather than an object enumeration task, it would suggest the deployment of separate memory retrieval and recognition mechanisms. Examining how observers enumerate displays with complex patterns, such as overlapping sets, would further develop this approach. In a study by Halberda, Sires, & Feigenson (2006), observers were asked to enumerate several
overlapping groups of dots that were differentiated into groups by their color. The primary finding is that small sets can be enumerated simultaneously even when these sets were interleaved among each other. Here, observers viewed a display for 500 ms that contained between 1-35 discs with 1-6 different colors. After the display presentation, observers were probed to report the number of discs of a specific color or of the entire display. Observers were accurate in reporting one set’s numerosity in this probe-after design when up to three sets of different colored discs were presented simultaneously. When more than three colors were presented on the display, however, errors increased by almost 100%. Probing observers before the stimulus presentation, however, did not increase the error rate when more than three sets of colors were presented. These results indicate that even non-spatially distinct sets can act as a single object representation in visual working memory (i.e., “single object” in the sense that a set is treated as one summary representation). This provides insight to the hierarchical representation of a scene in terms of treating items as individuals or as a set of individuals. These results, however, may rely on the opportunity to deploy serial attention and make eye movements to different regions of the display given the 500-ms presentation durations. Therefore, the results emphasize the ability of subitizing sets “serially” and maintaining working memory representations of these distinct sets in parallel.

Feigenson (2008) further explored the nature of parallel enumeration using a paradigm modified from infant studies: adults were probed on the number of objects that they saw being placed in different opaque buckets by the experimenter, while verbally shadowing an auditory recording of random letter sequences in order to prevent verbal counting. Observers were able to accurately report the numerosity only when there were
less than four types of objects (e.g., three sets) and overall failure occurred with four or more object types. This replicates the set-based limit observed by Halberda et. al (2006) and shows that a spontaneous enumeration of multiple sets occurs in parallel even with sequential and intermixed presentations of items in a set. Again, capacity was limited by working memory, where three simultaneous nonverbal numerosity estimates can be maintained and updated. Although the long stimuli presentations cannot attribute the enumeration of parallel sets to an independent subitizing mechanism, the results do indicate an interaction between enumeration and working memory representations (e.g., object files). Sets may be enumerated serially and transferred to working memory, where they are held and updated in parallel. Additionally, these studies show that not all enumeration is a pattern-retrieval process, since complex spatially-interleaved patterns can be enumerated accurately.

A novel account of enumeration is the proposal that the numerosity of a visual stimulus is an independent feature, like color, that can be detected by a dedicated mechanism in the visual system. Burr & Ross (2008) used the principle of sensory adaptation to test this hypothesis. Since visual properties, such as shape and color, are susceptible to adaptation and extended exposure can produce aftereffects, observers in this study were adapted to dot displays of varying numerosities to test if numerical estimation is a primary and independent visual property. Adapting to large numerosities (e.g., 400 dots) produced a decrease in perceived numerosity, but enumeration performance was more accurate for stimuli of 12 dots or less and at ceiling in the subitizing range. This effect was reversed when observers adapted to smaller numerosities (e.g., 50 dots), where overestimation occurred. These results were observed
even when controlling for texture density (by varying the size of the adaptor and test dots to vary pixel density) and when controlling for contrast (by varying the contrast level in the adaptor dots). The authors argue that the main factor that affected adaptation was numerosity and this indicates a distinct type of *qualia*—or phenomenological experience—for numerosity. This study also provides a single-mechanism account for enumeration even though small and large numerosities can be affected differently. Although this study presents an interesting explanation for the ability to detect numerosity, further studies are required to support this claim conclusively.

Overall, the results from studies examining the effect of features on enumeration support the view that strategies for enumeration can shift to suit the nature of the stimulus and the task. When presented with few targets, subitizing may occur because of a recognition of canonical patterns. As the number of targets increases to more than three items, a linear increase in RT may be due to “counting with attention” or from grouping processes. Since the maintenance of this count in memory is limited, larger sets require an estimation strategy characterized by responses with constant RTs but with increasing enumeration errors. This account of shifting strategies is feasible but needs more support to identify when and how such shifts are implemented.

### 2.2.2. Attention and enumeration

There are several forms of attention that can be recruited for enumeration. For example, counting larger sets requires spatial shifts in attention in order to individuate each item in the set, while subitizing or estimation can occur under global attention. Cueing regions in the visual field in order to focus attention on a distinct spatial region hinders the enumeration of displays with more than five items but leaves subitizing
unaffected even when these items are outside of focal attention (Trick & Pylyshyn, 1994a). This result may be supported by a rough spatial attention that can be deployed prior to feature-based attention (Liu, Stevens, & Carrasco, 2007). In such cases, the role of the visual indexing mechanism is suggested.

Some studies, however, have found subitizing to be sensitive to attentional load. For example, manipulating attention during an enumeration task via the inattentional blindness paradigm produces counting errors in all set sizes. In a study by Railo, Koivisto, Revonsuo, & Hannula (2008), observers were asked to judge which arm of a small cross was longest when it appeared briefly at the center of a screen—a feature-detection task that needed focused attention on a central location. When a random number of dots were flashed unexpectedly on the same screen, observers were asked to report that number. The first presentation of dots was the critical “inattention condition”, as observers had no instruction or expectation about the task. Subsequent presentations of the dots were treated as the “divided-attention” condition, and trials with only the enumeration task were treated as the “full-attention” condition. The results showed accurate enumeration in the inattention condition only if the enumeration display contained 1-2 items. This accuracy increased as more attentional resources were made available for the enumeration task in the latter two conditions. The results are taken to indicate that subitizing requires attention, since it was hurt when attention was focused on the primary feature-detection task.

In a similar study, Vetter, Butterworth, & Bahrami (2008) devised a dual-task experiment to test attentional load and its effect on enumeration. Here, enumeration was performed in addition to a feature-detection task with varying attentional demands. The
feature-detection task required the identification of a center diamond (comprised of four triangles) that was presented on the fovea. In the low-load condition, observers simply reported when a target color was present within the diamond. In the high-load condition, observers had to report when the diamond contained a specific spatial configuration of the target color—a feature conjunction. The enumeration component of the task required a simultaneous judgment of the number of target objects (1-8 vertically-oriented Gabor patches) among distractor objects (horizontally-oriented Gabor patches). Up to 13 Gabor patches could be presented in a ring equidistant from the center diamond (200-ms display). The enumeration results indicated that observers were more likely to overestimate numerosity in the high-load condition and underestimate when presented with four or more objects. As a measure of discriminability, the Weber fraction was computed and found to be larger in the dual-task conditions and especially large in the high-load conditions. The errors were greater in the subitizing range of both dual-task conditions and taken as evidence against a preattentive mechanism being responsible for subitizing.

Several elements of these studies, however, can be criticized. The poor enumeration found by Railo et al. (2008) in the “inattention” condition does not necessarily indicate that the (unexpected) count stimuli were not individuated; rather, they may have been individuated but ignored because the task did not require them to be processed (especially since they were easy to identify in subsequent “divided attention” trials). The enumeration task by Vetter et al. (2008) required the discrimination of targets and distractors among 13 objects prior to enumeration, which in itself taxes attention. Their conclusion that attention was required for subitizing may reflect, rather, that attention was required for searching through subsets in order to identify to-be-counted
targets, which is just one component of the task. Also, if subitizing is preattentive, then the center target in this experiment might be included automatically by the indexing mechanism and could account for the observed overestimation errors in the high-load condition. This interpretation could also explain why enumeration in both low- and high-load conditions was affected by the dual task, with a greater effect on the subitizing range. Attention is needed to transfer information about indexed items, and the observed errors in these studies may arise during this information-processing stage. These results do not address the preattentive nature of the individuation stage of this process, but rather the access to the items selected in this stage.

Studying the attentional blink is another way to test whether subitizing requires attention. The attentional blink is characterized by a decreased ability to notice a second target event that follows the first target event after a refractory period (Shapiro, Arnell, & Raymond, 1997). This visual “blackout” suggests that attention is required for processing the first target and cannot be deployed on a second target if it occurs within 150-500 ms after the first target. Egeth, Leonard, & Palomares (2008) tested subitizing during the attentional blink by using the rapid serial visual presentation (RSVP) paradigm. In a stream of single letters, observers were asked to report a red target letter and the number of green dots that flashed in a ring (either on nontarget or target letter displays). Enumeration accuracy was reduced, even within the subitizing range, when the dot display occurred within the attentional blink after the red target letter appearance. Another study by Olivers & Watson (2008) used a similar design to test if subitizing required attention. In this RSVP study, observers identified a target letter and enumerated a number of dots on a subsequent display. When the enumeration display fell within the
attentional blink, errors increased substantially for displays with two or more objects. The attentional blink also was induced when the task order was reversed. Attention may be important for enumerating all numerosity ranges and this conclusion is supported by similarly designed neuroimaging studies (Xu & Liu, 2008) and estimation studies (Burr, Turi, & Anobile, 2010). These results, however, are also subject to the criticism above: the performance decrement may be due to interference in the unconscious mental counting stage and not the individuation of items for enumeration.

These and various other studies claim to provide evidence against a parallel process with a fixed capacity that could be explained by an indexing mechanism. These claims, however, do not consider the possibility that indexing may still be parallel and preattentive but the access to the indexes for the purpose of deriving their cardinality requires serial attention. It is likely that this stage after preattentive indexing is subject to interference from the attentional blink, especially if the labeling of indexes is a sequential process. In fact, many recent studies overlook the necessity of attention to access indexes to report numerosity, as described in the original proposal by Trick & Pylyshyn (1989, 1993, 1994a, 1994b). Subitizing is said to occur before the allocation of serial attention but after more basic preattentive processes, like pop-out feature detection or grouping (Trick & Pylyshyn, 1993). An important distinction made in this preattentive proposal is that although subitized items may be preattentively indexed, attention is needed in a subsequent serial stage to assign a cardinal numeron. Since this information is assumed to be obtained by a rapid counting of the active indexes (that correspond to one individuated object), the RT rate has a shallow slope—it takes only slightly longer to recognize each additional item in the 1-4 range, increasing by 10-100 ms per item (Trick & Pylyshyn,
1994b). If this view is correct, it would provide another interpretation of the attentional blink results where the performance decrement may be due to a higher stage of processing, such as when numerosity labels are assigned or during the transfer of numerosity information into working memory, but does not affect the preattentive selection of individuals. Future studies should examine the possibility that a preattentive indexing mechanism used for enumeration is not affected by manipulations to attention, but instead interference occurs when transferring information about the active indexes.

In sum, many studies suggest that attention is important for enumerating sets of objects regardless of set size, but this result might be more indicative of the identification stage (i.e., assigning a number label) than the preattentive individuation stage. The conclusions made by some studies that subitizing is susceptible to manipulations of attentional load must be reexamined since the results may be due to interference from other processes, such as assigning a cardinality to active indexes or searching for target items. As it currently stands, the arguments against a preattentive indexing mechanism playing a role in subitizing are not conclusive.

2.2.3. Eye movements

When enumerating large sets of items, eye movements to different spatial regions is essential for serial counting and in the grouping of items to facilitate enumeration (Simon & Vaishnavi, 1996). Within the subitizing range, however, covert attention is sufficient for enumerating up to four items accurately even outside of the fovea (Trick & Pylyshyn, 1994a). Even when eye movements are prevented, enumeration remains rapid and accurate for small sets with average RT increases of only 25 ms per item (Klahr, 1973; Trick & Pylyshyn, 1994a), but more errors emerge when enumerating larger sets
This increase in errors when saccades are prevented (e.g., when counting in afterimages, Simon & Vaishnavi, 1996) suggests a recruitment of an additional process that is unnecessary for smaller sets. Eye movements also appear to play a role in the inhibition of previously counted items, which may explain the increase in saccade fixations when enumerating displays with larger numerosities (Peterson & Simon, 2000). One study that tested whether more fixations were required when enumerating displays with more than three targets among distractors found that the presence of distractors interfered with the ability to discriminate targets, and thus accurate enumeration required more fixations (Sophian & Crosby, 2008). This suggests that enumerating small sets is less dependent on attention (e.g., for saccade planning), but does not conclusively indicate a mechanism independent of attention. Instead, numerosity judgments may draw on both preattentive and attentional processes, depending on the spatial distribution of targets and the need to distinguish targets from nontargets.

Additional evidence supporting the need for saccades when counting is provided by a study that showed counting cannot occur in afterimages (Simon & Vaishnavi, 1996). Here, observers were shown dot arrays using a flashgun, which can induce afterimages that last up to 60 seconds. Enumeration was performed on the afterimages for 10- or 60-second durations. Enumeration errors were consistent outside the subitizing range and indicate that even with a generous amount of time to determine numerosity, observers needed to visually inspect the actual stimuli in order to selectively foveate individual items in larger dot arrays. Counting larger sets requires attending to different parts of a display more so than displays within the subitizing range. This supports the two-system perspective where small sets can be enumerated with covert attention, while displays with
more than four objects require eye movements in order to be counted serially. The need to keep track of already-counted items requires tagging processes (e.g., to prevent double counting), and eye movements may play a crucial role for this. A recent study, however, provides conflicting results where observers were able to enumerate from after-images in a seemingly unlimited way because the items were kept far enough apart and were easy to individuate in this study (Cavanagh & He, under review). This result suggests that the limit observed in the previous study could be attributed to crowding effects and emphasizes the importance of attentional mechanisms in subitizing (He, et al., 1997).

In general, eye-movements are unnecessary for accurate subitizing performance. This has been supported by studies that find the subitizing effect when restricting eye fixations to the central region (Watson, et al., 2007b) or using limited displays lasting under 200 ms (Mandler & Shebo, 1982), since eye-movements typically take around 200 ms to execute (Hutton, 2008). Counting larger sets, however, requires serial attention which often requires saccades to different locations. This is especially important for displays that are more difficult to parse (e.g., crowded displays) and when distractors need to be separated from targets prior to counting.

2.2.4. Working memory limits on enumeration

Working memory also plays an important role in enumeration, for example, in the verbal report of numerosity and the maintenance of multiple numerosity estimates. Trick (2005) showed that when a verbal counting task extends over the duration of sequentially presented items, it is more likely to interfere with working memory than a spatial enumeration task of simultaneously presented items. When demands on working memory are increased by requiring a secondary task (e.g., having observers tap their fingers while
performing the counting task), interference in enumeration only occurs for numerosities outside the subitizing range. This error is exacerbated when the secondary task requires complex articulation (e.g., vocalizing two different letters in sync with a metronome), suggesting that only counting is negatively impacted by an unrelated verbal task that requires working memory resources. Another study found a similar interaction of memory load on spatial selection: the number of spatial regions that could be selected simultaneously varied from 2-7 locations, depending on the required precision of the selection task (Franconeri, et al., 2007). This also supports a task-dependent notion of object selection that is influenced by the demands on resources that support information processing, which may be apparent in the attentional blink studies (i.e., with greater attentional demands, capacity for selection and therefore enumeration is limited). Other limits on working memory have been shown by studies on parallel enumeration, where multiple sets can be enumerated but limited to a maximum of three (Feigenson, 2008); this limit is thought to reflect working memory capacity.

Individual differences in working memory capacity affect distinct enumeration abilities (Tuholski, Engle, & Baylis, 2001). Observers who performed poorly on a counting task for larger numerosities also performed poorly on a working memory task (i.e., the “Operation Span” task, where observers performed an arithmetic problem while reading aloud a word). Within the subitizing range, however, the low working memory group had the same response latencies as high working memory observers. RTs in both groups were slowed when enumeration required processing a conjunction of features to separate targets from distractors, but the low working memory observers performed significantly worse in counting. The differences between the low and high memory span
groups suggest varying abilities in controlling inhibitory or attentional processes and working-memory encoding while counting. Such differences were not observed when enumerating four items or less. Although the reason for this is not clear, it may simply be that subitizing has low demands on working memory resources. Target identification, for example when searching through distractors, requires more working memory resources—it is this process that may be sensitive to manipulations of working memory.

These results emphasize the following question: is the source of errors during enumeration due to limitations in the transfer of information into working memory via selective attention? If so, this may leave intact the hypothesis that a preattentive mechanism is responsible for small-set enumeration. That is, preattentive subitizing is not affected by changes in attentional or working memory demands, but rather access to this information is disrupted, especially when verbal numerosity reports are required. The challenge lies in clearly identifying the stages in numerical processing, especially since it occurs on the order of milliseconds. The hierarchical nature of attentional processing adds additional complexity, as it may be possible to assign one index to a cluster of already-individuated items, for example, in the form a “set representation”. The relationship between indexes and working memory requires further study.

2.2.5. Language and enumeration

Although language plays an obvious role in verbal counting, its role in the development of enumeration skills and learning of numerical concepts is not clear. A strong Whorfian stance claims that language shapes number concepts. Therefore, only by learning a language can one form number labels and even understand mathematical principles. An alternative perspective claims that language does not provide core number
concepts but rather these concepts already are present as part of an innate “language of thought” that includes magnitude representations and a knowledge of mathematical principles such as addition (as reviewed by Gelman & Butterworth, 2005). This claim has been supported by studying cultures that have limited number vocabularies and studying children’s ability to reason numerically before learning count words.

Gallistel & Gelman (1992, 2000) argue that numerical reasoning is independent of language and emerges from nonverbal magnitude representations. When forming number words, labels are mapped to discrete increments on this continuous mental number line. The emergence of language-dependent skills, such as verbal counting, the ability to read or write numbers, and symbolic calculation using numerical notations, suggest an hierarchical architecture for numerosity representations that is supported by the nonverbal approximation system (Dehaene, 1992). From this perspective, the underlying development of numerical competency is access to a modular system of the analogical representation of numerical quantities, which is observed in preverbal children and animals. Although language provides symbolic representations of numerosity, numerical concepts are assumed to be independent of language and preconditions for learning linguistic labels (Dehaene & Mehler, 1992; Gelman & Gallistel, 2004).

Alternatively, the source for integer concepts and count lists may be a core system that provides a mechanism for individuating several objects in parallel and for referring to individual objects, such as the object file or indexing system (Carey, 2001). Since an object representation stands in one-to-one correspondence with the object it represents, Carey (2004) argues that number is implicitly part of an object file. The count list is first learned without any magnitude mappings—these symbols act as “placeholders” for
assembling richer meanings via mappings between systems of representations. The larger number concepts are formed via a process that involves directly learning verbal symbols and their relationships to quantity. This process draws on the human capacity for analogical reasoning and inductive inference that forms concepts in general (Margolis & Laurence, 2008). On this view, individual concepts are also considered units of thought that are individuated by their reference to entities in the world and their role in systems of inferential relations. In other words, the mapping between a cardinal label and the numerosity it represents must be learned like the relation between any verbal label and its concept during language development. Then, additional “bootstrapping” processes can form complex number concepts, such as fractions and irrational numbers (Carey, 2004)

How to explain the use of number words as young children mature continues to be debated. During language development, learning labels for category types and tokens occurs by exposure to their referents, which enables an association to be established between an object and its label. Since number need not refer to specific objects and can be generalized to any object category, it is problematic to apply such category-learning principles to number words. A more likely explanation is that a continuous magnitude is associated with a discrete count (as proposed by the accumulator model), and learning number words may depend on associating a label to a specific magnitude. Although it is unclear how this mapping of number words to a nonverbal magnitude occurs, it is clear that numerical reasoning and language develop independently.

2.2.6. Cognitive development and enumeration skills

2 A possible mechanism for this connection is discussed in Pylyshyn (2007) and Fodor (2009) in terms of visual indexes.
Developmental studies provide a rich ontogenetic survey of enumeration abilities from infancy into adolescence. It is well-documented that children reason numerically before exhibiting verbal counting skills (Barth, La Mont, Lipton, & Spelke, 2005; Gelman & Gallistel, 1978; Starkey, Spelke, & Gelman, 1990; Wynn, 1992; Xu & Spelke, 2000). These preverbal numerical abilities rely on mental magnitudes that can be used to add and subtract numerosities prior to attaining formal mathematical principles (Gallistel & Gelman, 1992). Continuous representations may produce ordered integer concepts with help from a generative mechanism that defines “next magnitude”, thus implicating an innate concept of “one” as well as a concept of successor or “next” (Leslie, Gelman, & Gallistel, 2008). Therefore, cardinality is thought to emerge early in childhood prior to the ability to use words to verbally represent count lists (Gelman, 2006) and may benefit from an innate ability to discriminate small sets of discrete objects (Starkey, et al., 1990).

In support of an early knowledge of cardinality, Zur & Gelman (2004) found that three-year-olds can predict the result of adding or subtracting 1-3 items to another quantity and then check their predictions by counting. Learning was also observed: if the child was exposed to a variant of an operation (e.g., \( x + y \)), then they could predict the inverse \( (x - y) \). Other studies suggest that two-year-olds can identify sets of 1-3 prior to being able to apply the successor function (i.e., adding or subtracting) on that numerosity, suggesting that cardinality must be understood for performing arithmetic but is unnecessary for detecting exact equality (Sarnecka & Carey, 2008). An understanding of exact equality also is evident in cultures with a limited lexicon of number words. For example, the Mundurucú, an Amazonian indigene group, have a limited understanding of the successor function and their language does not have exact numerical symbols, but
they still grasp the concept of exact equality (Izard, Pica, Spelke, & Dehaene, 2008). Further studies on the successor function may clarify how numerical abilities develop.

Studies on infants’ ability to discriminate between sets of objects support the presence of both continuous and discrete number representation systems that are engaged according to task demands (Cordes & Brannon, 2008b; Feigenson, Carey, & Spelke, 2002). One of these systems (for magnitude representations) allows infants to determine continuous values such as surface area, contour length, and ratios of numerosity. Continuous representations are observed when discriminability between sets of items follow ratio-dependent patterns or a logarithmic scale. For example, studies using large numerosity displays indicate that a twofold increase in set size is required for six-month-old infants to notice that one set is larger than the other. Infants notice changes in set size when comparing 8 vs. 16 objects but not 8 vs. 12 (Xu, Spelke, & Goddard, 2005). Similarly, infants notice a fourfold increase in area regardless of set size, but cannot notice a threefold increase (Cordes & Brannon, 2008a). Another study tested six-month-old infants on their ability to discriminate changes in the size ratio of two objects by changing the surface area of cartoon faces (Brannon, Lutz, & Cordes, 2006). Here, infants did not notice a change when the ratio of surface area increased by 50% but did so after a twofold increase in the ratio. By ten months, however, infants are able to discriminate between a 2:3 ratio (e.g., 8 vs. 12 items) (Xu & Arriaga, 2007). Such studies support the use of magnitude representations with ratio-dependent discriminability that increases in acuity over the course of early childhood.

A second system for number representations may rely on the innate parallel individuation system that contributes to the perception of the cardinality of a set in a
manner distinct from the detection of numerosity by counting or approximation (Carey, 2001; Feigenson, et al., 2004; Klahr, 1973; Silverman & Rose, 1980; Starkey & Cooper, 1980; Svenson & Sjoberg, 1978; Trick, Audet, & Dales, 2003). In support of this view, one study found that three-year-olds can determine numerosity more accurately when an array of dots were presented simultaneously rather than sequentially (Benoit, Lehalle, & Jouen, 2004). This indicates that a parallel mechanism helps to enumerate small sets; if a serial counting was employed, it should not have been disrupted in the sequential presentation condition. In fact, a child’s ability to attend to numerosity spontaneously has been directly related to the ability to notice a number of items in parallel and this skill increases in capacity with practice and maturation (Hannula, Räsänen, & Lehtinen, 2007). Subitizing ability reaches that of adults between the ages of three and five, and suggests a benefit from experience or maturation (Starkey & Cooper, 1995) that can continue to increase in capacity and speed with age (Trick, et al., 2003). For example, processing an additional object within the subitizing range takes 100 ms in seven-year-olds and 30 ms in adults, and counting time for larger sets decreases from 1000 ms to 250 ms per additional item (Svenson & Sjoberg, 1983). Results from a recent study also indicate that acuity in making magnitude estimates increases during childhood development, reaching the resolution level of adulthood by early adolescence (Halberda & Feigenson, 2008). This may rely on the maturation of the visual system and the attentional processes that interact with the individuation mechanism.

Evidence for a discrete quantitative system in infants has been inferred from their ability to represent precisely sets of 1-3 items, without ratio-dependent discriminability. In one study, a manual-search paradigm was used to test 12-month-old infants’ ability to
represent objects that were placed inside a box and out of view (Feigenson & Carey, 2003). When the infant saw three toys placed into the box and then saw the experimenter remove two, the infant searched longer for the third toy than in cases where four toys were placed in the box and two were taken out. The insensitivity to sets of four or more suggests an accurate representation of sets less than four (e.g., by a parallel individuation system). Another study tested 7-month-old infants’ preference for attending to continuous or discrete properties by manipulating the set sizes which were displayed and observing whether the infants noticed a change in number (Cordes & Brannon, 2009). This study used small sets (2, 3), larger sets (8, 16), and both (2, 8) and also manipulated continuous properties such as total contour length and area occupied by the shapes. The results indicate that infants can attend to a discrete number of items when continuous properties were held constant and when they varied in opposition to numerosity. These results challenge previous findings that infants prefer continuous representations (e.g., Feigenson, et al., 2002).

Even five-month-old infants can efficiently represent the number of collective sets in a visual display (Wynn, Bloom, & Chiang, 2002). When groups of dots move randomly but in a manner that creates a distinct set of objects, the sets are treated as individuals and infants can detect a change in the number of these collections. Another study found a similar ability to group moving items into sets and thus increase the representational capacity of number (Feigenson & Halberda, 2004). A related ability that increases during childhood is information “chunking” in working memory. One study examined auditory chunking abilities in first-graders, sixth-graders, and adults (2009). The results showed a developmental increase in the number of chunks that could be
recalled, but with no increase in the amount of information that a chunk could contain. Furthermore, the ability to enumerate can increase with normal childhood development and become enhanced through spatially demanding tasks. In fact, even adults can exhibit an increased subitizing range, which is attributed to changes in visual short-term memory skills from playing action-based video games (Green & Bavelier, 2006).

Stimulus properties, such as color or shape, seem to have little effect on counting performance. A study by Gelman & Tucker (1975) asked children (ages 3-5) to choose the display that contained more items. When asked to choose which of two plates had more items, the children chose the correct group even if the items within the group changed identity (e.g., homogeneous groups became heterogeneous). The results indicate that young children can ignore heterogeneity and focus on individual items by treating color or identity substitutions as number-irrelevant transformations. This result is supported by the finding that object properties can be used to infer how many items were placed behind a screen, but not notice a change in features (Tremoulet, et al., 2000). Other studies have shown that four-year-olds can count heterogeneous displays with homogenous subgroups better than completely homogenous displays, possibly because it makes it easier to keep track of what has been counted (Schaeffer, Eggleston, & Scott, 1974). This bias towards counting by token type is not seen in older children. It is possible that only when a child becomes well-versed in counting through practice that they develop strategies that allows them to skip the serial counting process and simply chunk the array to be counted (Gelman & Gallistel, 1978, p. 70).

Abnormal development also characterizes the mechanisms of numerosity representations. In cerebral-palsied children, subitizing capacity is reduced from damage
to the right brain hemisphere that affects visuospatial processing (Arp & Fagard, 2005). Poor enumeration of small sets may be due to the inability to perceive and remember spatial configurations. There is also a strong relationship between subitizing limits, counting ability, and hand-eye coordination, where hand-eye coordination accuracy was a reliable predictor of subitizing limits (Arp, Taranne, & Fagard, 2006). This may indicate the dependence of subitizing on a mechanism that enables acting upon visual objects, such as object or event files (Hommel, 2004). Finally, poor subitizing performance appears to be indicative of mathematical learning disabilities and can be detected reliably by age five (Desoete & Grégoire, 2006; Halberda, Mazzocco, & Feigenson, 2008).

The overall results from developmental studies indicate that preverbal infants can represent numerosity via magnitudes independent of language skills and can perform basic mathematical operations on these magnitude representations. Additionally, the ability to discriminate distinct visual objects is present during infancy becomes more efficient during development, which may explain why enumeration capacity increases. Language also may supplement these abilities by providing symbols that serve as a form of external memory.

2.2.7. Aging effects on enumeration

Enumeration abilities change over the lifespan. Older adults show slower counting but subitizing remains unaffected (Trick, Enns, & Brodeur, 1996). In normal aging, subitizing speed tends to be the same as young adults, although searching through distractors that are similar to the targets becomes slower (Sliwinski, 1997; Watson, Maylor, Allen, & Bruce, 2007a; Watson, Maylor, & Bruce, 2005a). The ability to track multiple objects simultaneously also declines from four to about three items (Trick, Perl,
& Sethi, 2005). These performance changes suggest that high-attention tasks, such as those requiring the inhibition of distractors, are impacted by the aging process, but low-attention tasks, including subitizing, continue to operate efficiently.

Aging may shrink the breadth of attention focus but may not affect the ability to access the items in the focus of attention (Basak & Verhaeghen, 2003). In a study using a Stroop-like counting task, observers enumerated a set of digits where the numerosity value of the digit used could differ by +/-1 from the total number of digits on the display. The Stroop effect, where the numerosity of the set of digits did not correspond to the value of the digit used, was observed in the counting range but not in the subitizing range for both younger and older adults, which supports an automatic account of subitizing. Older adults, however, exhibited a smaller subitizing range than the younger adults, which may be due to the higher attentional demands of Stroop-like tasks. An observer’s subitizing range may be taken as an indicator of the degree of the focus of attention (Cowan, 2001), and the observed decrease in the subitizing range may be due to shrinking attention capacity (for accessing individuated items) or a decline in the inhibitory control of attention (Sweeney, Rosano, Berman, & Luna, 2001). Another explanation for the slowing of numerical processing in older adults may be due to age-related changes in the ability to plan and execute eye movements, which is required for searching and serially counting through larger sets of items, but this may be specific only to people with certain forms of dementia (Watson, Maylor, & Bruce, 2005c).

Performance on enumeration tasks also can help identify abnormal aging. For example, adults with Alzheimer’s show an intact ability to enumerate but may forget arithmetic facts and other procedural rules (Kaufmann, et al., 2002). This suggests
independent forms of numerical abilities and makes it less likely that subitizing relies on a memory-related recognition of patterns. Enumeration, however, is significantly slower in patients with Alzheimer’s, who also exhibit a reduction in the subitizing range from 3.5 to 2.3 items (Maylor, Watson, & Muller, 2005). This slower enumeration has been observed in the early stages of Alzheimer’s, before the subitizing capacity reduction (Maylor, Sheehan, Watson, & Henderson, 2008). Such enumeration tasks can prove to be useful in diagnosing cognitive dysfunction, such as Alzheimer’s and vascular dementia (though distinguishing between the two is not possible with this task).

As people age, subitizing is less affected than counting or other processes that require higher attentional resources, which supports a preattentive account of subitizing. Such results can help distinguish which stages in the enumeration process are preattentive and which ones require attention or working memory resources. Namely, the individuation process, which supports subitizing, does not require the same amount of cognitive resources required to access or assign a label to the individuated objects.

2.2.8. Insights from neuroscience

Brain imaging techniques have been used to investigate the neural substrates of numerical processing, such as the mapping of symbolic representations to nonverbal numerosity estimates (Lyons & Ansari, 2009) and characteristics of the mental number line (Dehaene, 2003; Nieder & Dehaene, 2009). Attempts to discern whether enumeration of small and large magnitudes involves two neural systems indicate that either a parallel or serial process can be engaged for subitizing, depending on the complexity of the stimulus (Piazza, Giacomini, Le Bihan, & Dehaene, 2003). Some studies measuring brain activity when counting found no distinct neural systems for small or large numerosities
per se, but find more activation during the counting of larger sets in regions of the brain involved in shifting spatial attention (Piazza, et al., 2003; Piazza, Mechelli, Butterworth, & Price, 2002). This increase in activation is directly related to the difficulty of the counting task and may be due to the tagging of already-counted items.

Examining the brain regions associated with attention during enumeration may clarify its role in subitizing. Using an attentional blink paradigm, Xu & Liu (2008) manipulated attentional demands and measured event-related potentials (ERPs) to identify brain regions engaged during enumeration. In some trials, observers only reported the target letter in a RSVP task; in dual-task trials, they also reported the numerosity of a display shown after the target letter, which could appear within or outside of the attentional blink. All enumeration suffered in the dual-task condition when these displays appeared within the attentional blink and affected brain regions responsible for modulating attention. This result suggests that attention is required for identifying numerosity. The reduced availability of attentional resources during the dual-task trials corresponded to the behavioral results of decreased enumeration accuracy. These results also may be due to a growing resource demand that increases with each additional element to be counted (Balakrishnan & Ashby, 1992). On the other hand, these results could simply indicate that processing items presented in RSVP tasks requires more resources or reflects the limitations in other processes (as discussed earlier)—the current results are inconclusive.

The strongest support for the engagement of multiple systems for enumeration is found in studies that indicate subitizing (and other low-level visual processing) is localized in areas of the occipital cortex but serial counting activates additional areas in
cortical and subcortical regions (Sathian, et al., 1999). This suggests that counting and subitizing use separate systems, or at least different combinations of systems, which may be related to the employment of a serial process that shifts attention spatially for counting. A recent fMRI study supports this hypothesis, where activation for displays with numerosities within the subitizing range activate a more “stimulus-driven” region of the brain (the right tempero-parietal junction), which and was suppressed during the “goal-driven” processing of larger numerosities (Ansari, Lyons, van Eimeren, & Xu, 2007). This distinction between stimulus-driven and goal-driven processing of numerosity displays supports a multiple-system account for enumeration.

The hemispheric specialization of numerical processing also may identify the systems that support enumeration. A study manipulating the presentation of stimuli across visual eye fields identified a right-hemisphere preference for processing numerosity displays within the subitizing range while the counting of larger sets was better processed by the left hemisphere (Pasini & Tessari, 2001). Neuroimaging studies also report this specialization and support the involvement of two distinct neural systems for different enumeration strategies, such as verbal or nonverbal counting (Gandini, Lemaire, Anton, & Nazarian, 2008). Other studies have identified consistencies in brain mechanisms across individuals that may account for the increased capacity for numerical processing, with evidence supporting the intraparietal sulcus within the parietal cortex for numerosity detection (Cantlon, Platt, & Brannon, 2009), which can be identified in infants as young as three months (Izard, Dehaene-Lambertz, & Dehaene, 2008).

Overall, the neuroscience literature support a two-system account for enumeration, suggesting a flexible shifting of strategies based on numerosity or task...
demands. If this is the case, there may be two types of systems available for subitizing that are employed according to the complexity of the task and the availability of cognitive resources such as attention.

### 2.2.9. Concluding remarks on enumeration

The vast literature on numerical processing underscores its central role in cognition—from estimation, to counting, to complex mathematical abilities. This review touches on a subset of the studies examining numerical cognition, with a focus on whether small-set enumeration may involve a distinct “subitizing” mechanism. There is evidence that multiple systems can be involved in enumeration, depending upon the task demands. The magnitude representation system enables the efficient estimation of numerosity, as described by the accumulator and mental number line models. This mechanism can serve all enumeration and is likely present in studies that observe scalar variability in performance. In fact, this application of Weber’s Law is often used as a benchmark for the presence of an underlying continuous magnitude representation. Other support for a single enumeration mechanism comes from studies that claim that attentional manipulations affect both subitizing and counting, suggesting a common representational system for all number ranges. It is possible, however, that these results do not reflect the subitizing mechanism but other information processing limitations. In other words, the source of noise in human performance may originate from other cognitive processes, such as visual search (for processing targets from nontargets) or the transfer of information into working memory. The single-system arguments that rest on the results from studies using attentional manipulations require further support, since the
source of these results remains questionable (i.e., the stage during enumeration that is affected).

Other cognitive systems that may be engaged during small-set enumeration include a pattern-matching system for familiar stimuli or an object individuation and indexing system that can select in parallel a limited number of distinct visual objects. Furthermore, an individuation mechanism may facilitate the ability to map discrete objects to discrete increments on the mental magnitude representation (e.g., was well as help keep already-counted and to-be-counted items distinct). Given the importance of numerical cognition, it is reasonable to expect that the human cognitive system has evolved to employ several methods to obtain numerosity information. Therefore, in addition to a primary magnitude representation system, it is likely that an object individuation system may be used in small-set enumeration. The challenge, however, lies in clearly identifying the mechanisms involved and in determining the conditions under which they are employed. To this end, the role of the visual indexing mechanism in subitizing will be examined further in the current study.
Chapter 3. Study Goals

The goal of the current study is to examine how spatial attention interacts with the ability to enumerate sets of objects. Although both topics have been studied extensively, the role of spatial information in subitizing has yet to be examined closely. Judging quantity is a basic cognitive function that is supported by a primitive magnitude estimation mechanism and likely augmented by an object individuation mechanism. In order to understand how these different systems may be implemented, it may be useful to examine the role of spatial information during enumeration. If the encoding of spatial information is required for successful enumeration, then the system involved must be linked to higher-level processing that can encode this information, such as an object file system that stores information about perceived objects. If location accuracy is not crucial for reporting accurate numerosity, then a preattentive mechanism, such as the indexing mechanism that can individuate and provide a reference to individuals without necessarily encoding any of their properties, may be all that is needed for enumeration. According to the Visual Index (FINST) theory, these indexes are automatically assigned to distinct objects without making use of information (if any) that may be stored in their object file representations, including location information. Furthermore, the indexing view claims that in deriving the cardinality of the set of objects, the only things that need to be enumerated are the active indexes and not the objects in the scene, which have already been individuated and indexed. This allows the process to be fast and accurate.

The current study introduces a new experimental methodology designed to examine the relationship between the representation of sets of objects selected for enumeration and the representation of their spatial properties. The experimental task
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. During this task, observers were shown a brief stimulus (50, 200, or 350 ms) comprised of 2-9 small black discs randomly placed on a gray screen, which was immediately followed by a mask to ensure a controlled processing duration of the stimulus displays. These stimuli were presented briefly in order to prevent verbal counting and limit object file formation. Then the observers used the computer mouse to “point” to where the objects had been by placing markers on a blank screen at the former locations of each disc. This response method allowed for a nonverbal report of numerosity and location under the assumption that being able to report on a number of locations indicates the observer’s capacity for individuating a number of items for enumeration; thus, this subitizing measure precedes the assignment of a numeral label. This methodology allowed us to analyze performance for both enumerating and localizing the stimulus displays, and to examine the relationship between these two measures. *Enumeration accuracy* is measured in two ways: as the percent of trials correctly enumerated and as the average number of miscounts. For each trial, each response marker was paired with a disc on the stimulus display (in order to determine which response marker most likely corresponded to a disc on the stimulus display); *location accuracy* is the distance between these corresponding discs.

By presenting identical objects, there should be little feature competition for which objects are selected for enumeration. The displays were designed to prevent

---

3 Note on terminology: The number of items recalled was measured by counting the number of item-locations observers marked. This is taken as the number of items “enumerated” 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 (even though the observer does not actually provide a cardinal numeral response).
crowding (therefore easily discriminable) and did not extend more than 10.5° from either side of the central fixation. This allowed examining the limits of enumeration under more “optimal” conditions. This will allow us to describe how other spatial properties affect enumeration performance, such as proximity to fixation at time of presentation. The spatial relation between the stimulus discs was recorded and analyzed to determine if certain characteristics, such as regularity in spacing, affected task performance. The spatial relations among the objects on the response screen, which were placed by the observers, also were analyzed and compared to stimulus displays. Since systematic localization errors have been observed in studies on spatial memory, such effects can be examined in the current results. For example, previous studies have identified a bias in location memory toward geometric “prototypes” (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Newcombe, & Sandberg, 1994), and these biases increased as memory became less certain over extended response delays (Spencer & Hund, 2002). The spatial configuration of the response markers may produce results to support these observations.

The manipulation of the presentation duration allowed testing the time-course of item selection (i.e., indexing) and the encoding of spatial information. The briefest presentation duration (50 ms) might prevent serial attention from visiting each object. The longer presentation durations should allow better encoding of properties, such as locations, since there is more time for attentional processing. If this distinct feature encoding stage is maintained, it could indicate that subitizing occurs prior to higher-level processes (e.g., object files), supporting a preattentive hypothesis of subitizing.

---

4 The most direct estimate of attention-scanning may be one by Tsal (1983) at 117 degrees/sec. Under this assumption, the total distance attention could travel at 50 ms would be about 5.8 degrees, not counting the time to encode locations of objects along the way.
Furthermore, shorter stimulus durations may extenuate the spatial errors typically observed in spatial memory because of the limited encoding time.

The use of the “pointing” method measures “implicit” numerosity and provides evidence that observers process significantly more items than they are able to indicate when reporting the numerosity of a set with a cardinal label. Since this is a novel finding, a number of follow-up experiments were designed to determine which aspect of the pointing method is primarily responsible for the increased number of items detected. The overall results support an individuation mechanism in early vision that facilitates a nonverbal enumeration and also can represent the spatial configuration of a set of visually distinct items. This process can select and report on sets containing up to six items with high accuracy even when the stimuli are presented very briefly (e.g., 50-ms). Localization accuracy, however, was sensitive to temporal constraints and increased in precision with longer viewing durations (i.e., from 50-ms to 200-ms displays). Additionally, these processes were sensitive to the level of spatial regularity on the stimulus displays. Although these findings are preliminary and require further empirical support, they do provide a unique starting point for studying the individuation process and its interaction with the way in which people determine numerosity.

By studying subitizing, we hope to learn more about the limits of visual information processing and how attention interacts with early vision. Visual Indexing Theory describes a mechanism that explains the ability to select four or five objects rapidly and accurately, even under very short durations. Since visual indexes are thought to operate early in attention (i.e., at a “preattentive” stage), comparing observers’ performance on reporting numerosity and their performance on localization accuracy
should help clarify the stages of small-set enumeration. Additionally, characterizing localization errors based on scene properties (e.g., spatial configuration) can help illustrate how the indexing mechanism operates under various conditions.
Chapter 4. Experimental Methods & Results

4.1. General Methods

The same design for the “enumerating-by-pointing” task was used in all the experiments reported and is described below. The control conditions introduced in Experiments 2-5 used the same stimuli and apparatus to collect data, but the response method varied in each Experiment. The reporting method for the control conditions are described in the corresponding sections.

Apparatus & Stimuli

These experiments were programmed in MATLAB® using Psychophysics Toolbox 3.0.8 (Brainard, 1997) and controlled by PC computers running the Windows® XP operating system. The stimuli were displayed on either a 21” color HP P1100 CRT monitor or 19” color Sony Trinitron® CRT monitor (both with 1280x1024 pixel resolution at 70 Hz). The monitor settings for contrast and brightness were set to 10% and 50% respectively to reduce phosphor decay lag.

The test stimulus consisted of identical black discs that appeared on a gray background (RGB values were [75, 75, 75]) optimized to reduce contrast and minimize after-images and phosphor decay. The discs were 35 pixels in diameter (~1° visual angle) and randomly placed on the screen with the following constraints: disc edges could not lie within 115 pixels (~3°) and no farther than 715 pixels (~20°) of each other; additionally, discs could not appear within ~200 pixels (~5°) from the screen edges. This produced an effective viewing display of 768x614 pixels, or approximately 21° by 17°. The minimum distance between discs was set at ~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).

**Procedure for the “Enumerating-by-Pointing” Task**

Observers sat approximately 60 cm from the screen in a darkened room. They were instructed to look carefully at 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 the designated duration (e.g., 50, 200, or 350 ms), followed by a 16-ms black screen and an 85-ms mask. The mask was generated by randomly assigning a white or black value to a grid of 4x4 pixel squares that covered the entire display. This display was designed to optimize masking by presenting a high-contrast mask over the low-contrast stimulus displays, and by introducing a blank screen during a short lag between the stimulus and mask (Hermens, Luksys, Gerstner, Herzog, & Ernst, 2008; Reeves, 2007; Sperling, 1965). After the mask, a gray screen appeared with a crosshair pointer at the center of the screen. Observers placed markers on the recalled location of each disc (see Figure 1). The marker was a small black “X” approximately 16 pixels in width (~0.5°). 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 observers finished marking the disc locations, they pressed the space bar to start the next trial. They were encouraged to take a break at any point during the experiment and especially during the midway point indicated by the experimental program (although most participants did not take a break).
Measures & Analyses

The primary measures of interest were enumeration accuracy and localization accuracy. *Enumeration accuracy* is examined in two ways: 1) as the proportion of trials in which the correct number of markers were placed on the screen, and 2) as average “miscounts”, which was the actual number of objects present in the display minus the number of markers placed (the absolute value). The average miscounts were transformed into absolute values to simplify reporting and avoid averaging errors that might inflate accuracy. Additionally, most counting errors were undercounts (on average, ~85% of errors were undercounts). *Localization accuracy* is reported as the Euclidean distance between the center of a stimulus disc and the center of a paired response marker (the method for pairing these items is described below).

The data were analyzed using PASW® Statistics 18 software package (SPSS Inc., 2010). Outlier responses were identified and excluded from any analyses; these were miscounts greater than 5 (< 0.1% of trials) and location errors greater than 500 pixels or 13.7° (< 0.3% of trials). Analyses of variance (ANOVA) were conducted on the measures using a within-subjects design and included subject ID as a random variable to account for between-subject variability and excluded the last numerosity condition (9 items) to reduce end effects. The magnitude of the effects were reported in terms of the *partial $\eta^2$* measure, which accounts for the variance from multiple factors (duration, numerosity, or reporting method); this is noted as $\eta^2_p$ in this paper.

---

5 These end effects could include an anchoring effect where performance increases in the last numerosity condition since observers may expect to respond a certain number of times in each condition. This was not the case in the current results, however; instead, extremely low performance was observed, which increased the observed effects of numerosity. Removing the largest numerosity produced a more conservative estimate of effect sizes for the numerosity factor.
In addition to using the Bonferroni test to adjust for multiple pairwise comparisons, the enumeration measures were analyzed using the Scheffé test, which is a conservative test of significance for multiple pairwise comparisons based on the $F$ distribution. An advantage of this method is that it tests all possible linear combinations of group means (not just pairwise comparisons) and identifies homogeneous subsets of group means (SPSS Inc., 2009). This can identify where the linear trends deviate from “flatness” (i.e., which group means are not statistically different) and may be used to identify a “subitizing range”. A trend analysis also was performed on the enumeration measures to identify deviations from linearity in the following numerosity ranges: 2-4, 2-5, 2-6, 2-7, 2-8, and 2-9. This sort of trend analysis has been used in previous studies to determine the subitizing range from RT data (e.g., Akin & Chase, 1978; Chi & Klahr, 1975; Trick & Pylyshyn, 1993), where the largest number range that does not deviate from linearity is taken to be the subitizing range. This assumes that the linearity corresponds to a slope near 0, and thus indicates a “flatness” in performance. The problem with this method is that it does not detect a true flatness per se, since a linear trend may possess a constant increase in errors; therefore, the deviation from linearity that is identified is a deviation from a linear rate of increase and not flatness. The slope of these errors is reported in order to indicate the average rate of increase in errors within the subitizing range identified by the trend analysis.

A more stringent test for flatness in enumeration performance was performed using Bayesian hypothesis testing. This test compares the null hypothesis (i.e., enumeration accuracy for sets with more than two items is equal to the accuracy of sets with two items) to an alternative hypothesis (i.e., accuracy for the higher number of items
is less than for two items). In formulating this alternative, one must specify a limit on the effect size, that is, on how much worse the latter performance might be. If one specifies an unrealistically large possible effect size, then the comparison may favor the null even though there is a clear but small effect. This is the “problem of the prior” in Bayesian comparisons. For this analysis, the solution proposed by Gallistel (2009) will be adopted, which is to compute the odds for/against the null as a function of the limit on the effect size. If there is no effect, then this function gives odds in favor of the null that decrease toward even odds (1:1) as the limit on the effect size decreases toward 0, at which point the alternative to the null is no longer an alternative. This happens, for example, when enumeration performance for the higher number displays is better or equal to that for the two-item displays. When this happens, the null is demonstrably unbeatable, given the data, and should clearly be accepted. If performance on the higher number displays is even a little smaller than for the two-item displays, then this function will dip below even odds when the limit on the effect size is just generous enough to encompass the likelihood function on the lower percentage. This likelihood function represents the uncertainty regarding the true value of performance. The maximum extent to which the odds turn against the null indicates how strongly the data would support an alternative to it of the form “there is some diminution in performance”. If the odds are at best no greater than 3:2 against the null, then one can say that there is very little reason to reject it. In other words, there is very little evidence against flatness. If the odds against the null exceed 3:1, then the data give some justification for rejecting flatness; if they exceed 10:1, then there is overwhelming justification for rejecting flatness (Jeffreys, 1961; Kass & Raftery, 1995). This analysis was performed on the proportion of trials enumerated
correctly as another method to identify the numerosity conditions in which performance was flat in order to provide a more conservative computation of the subitizing range.

Finally, a simple linear function was computed to compare the average numerosity response to the actual display numerosity. This illustrates response precision and was used to compare performance in different numerosity ranges, namely for a subitizing range of 2-4 versus a subitizing range of 2-6.

Processing the Location Data

The location data was processed in MATLAB® using built-in functions and toolboxes (The MathWorks, 2009). The location data consisted 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 items were paired based on their coordinates (x-y values) using the following procedure. When a trial had the same number of stimulus and response items (i.e., a correctly enumerated display), 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 reference set by factoring out variations in scaling, rotation, and translation (Goodall, 1991). In the observers’ localization data, this method can remove the “compression” that occurs, which reduces errors during the pairing procedure. After applying the relevant scaling, rotation, or coordinate position transformations on the response data, a nearest-neighbor method that used Delaunay Triangulation (Kendall, 1989) identified the likely associated response marker for each stimulus disc in each of the trials, including those not eligible for the Procrustes transformations (more details about triangulation is described in
Error was calculated as the pixel distance between the center of a stimulus disc to the center of the original response marker (i.e., not the transformed coordinates used to establish the pairs). Some trials resulted in unpaired discs, for example, in miscounted trials. These were excluded from the location analysis (approximately 78% of possible disc pairs).

Summary of Experiments 1-5

Experiment 1 presents data from the first full (non-pilot) implementation of the enumerating-by-pointing task described above. In addition to the pointing task, Experiments 2-5 included control conditions where observers also completed a block of trials using the same stimulus as the pointing task but with an alternate method for reporting numerosity. This allowed for a within-subjects comparison of results from the pointing task to alternate reporting methods. To compare the results of the pointing task to a more “conventional” reporting method, Experiments 2A and 2B required observers to select an Arabic numeral from a response display that appeared 85-ms after the mask. This display was either configured in a clock-like manner in Experiment 2A, or presented in a single row at the bottom of the screen in Experiment 2B (depicted in Figure 5). In Experiments 3A and 3B, the pointing task was compared to an alternate nonverbal reporting method. Here, observers recorded their numerosity response with a simple “tally” that provided a visual marker for each click they made, but in a spatially-irrelevant way. These tally marks appeared at the center of the screen in Experiment 3A, or at the bottom of the screen in Experiment 3B. To examine performance when making a simple clicking response, the control block in Experiment 4 collected the numerosity response by recording the number of times the observers clicked the mouse (without
continuously visible markers). Finally, to examine the role of visiting locations in making
the response, the control block in Experiment 5 used a response method almost identical
to the standard pointing task, except that the markers only appeared on the screen for
200-ms. These methods are described below.

Figure 1. Schematic of the enumerating-by-pointing task. The first panel depicts the initial screen
in each trial, with a central fixation cross. After 2,500 ms, a stimulus display with 2-9 randomly
placed black discs appears for 50, 200, or 350-ms. This is followed by a black screen for 16-ms
and a random-dot mask for 85-ms. The final screen depicts the response stage where observers
mark the locations of the stimulus discs using a mouse (depicted by the white crosshair pointer).

4.2. Experiment 1 – Enumerating-by-Pointing Task

Experiment 1 is the first full experiment to examine the enumerating-by-pointing
method. The enumeration stimuli were generated as described above and were presented
for 50, 200, or 350-ms durations. This allowed us to examine how enumeration and
localization performance was affected by changes in stimulus duration and set size. These
results were reported in Haladjian & Pylyshyn (2011).
Method

Experiment 1 uses the method described above. There were three display durations of 50, 200, or 350 ms and eight numerosity conditions of 2-9 discs. Observers received 12 trials of each of the 24 possible test conditions (3 durations × 8 numerosities); these 288 trials were randomly distributed throughout the experiment as determined by the experimental program. Observers were encouraged to take a break at any point during the session. Twenty-four Rutgers University undergraduates participated in one 45-minute session for course credit or payment of $10 or two movie tickets. (Six additional students completed the experiment but were excluded from the analyses due to errors in the timing of the stimulus displays due to programming errors.)

Results

Enumeration accuracy

Enumeration accuracy was measured as the proportion of trials in each condition in which the observer provided the exact correct number of location marks. This measure showed that performance was high for displays containing up to six items and decreased significantly for larger numerosities. The ANOVA results for proportion of correct trials indicates main effects for display duration ($F(2,5543) = 34.7, p < 0.001, \eta^2_p = 0.602$) and numerosity ($F(6,5543) = 68.8, p < 0.001, \eta^2_p = 0.750$), with an interaction ($F(12,5543) = 7.7, p < 0.001, \eta^2_p = 0.250$). Figure 2 shows the enumeration accuracy as a function of numerosity for each display duration and suggests that the performance in the 50-ms displays was worse than the longer display times. For the smaller numbers of items there was little difference, possibly due to a ceiling effect.
The average number of miscounts in each condition was analyzed. Over- and under-counts were treated the same in this analysis by using the absolute value of miscounts (83% of errors were undercounts). ANOVA results for miscounts indicate significant main effects for duration \((F(2,5541) = 41.2, p < 0.001, \eta^2_p = 0.642)\) and numerosity \((F(6,5541) = 56.9, p < 0.001, \eta^2_p = 0.712)\), with an interaction \((F(12,5541) = 10.8, p < 0.001, \eta^2_p = 0.319)\). The average counting error increased with greater numerosities, but less so for the longer durations (see Figure 3).

A Bonferroni test for multiple pairwise comparisons confirmed that the counting performance in the 50-ms condition was significantly worse than in the 200-ms condition for displays with 3 items (mean difference: \(MD = 0.03, p < 0.05\)), with 6 items \((MD = 0.09, p < 0.05)\), 7 items \((MD = 0.12, p < 0.01)\), 8 items \((MD = 0.26, p < 0.01)\), and 9 items \((MD = 0.46, p < 0.01)\). The differences between the 50-ms and 350-ms conditions were significant for displays with 3 items \((MD = 0.03, p < 0.01)\), 6 items \((MD = 0.12, p < 0.01)\), 7 items \((MD = 0.20, p < 0.01)\), 8 items \((MD = 0.36, p < 0.01)\), and 9 items \((MD = 0.60, p < 0.01)\). The only significant difference between the 200-ms and 350-ms conditions was observed in displays with 9 items \((MD = 0.14, p < 0.05)\).

**Subitizing range**

The Scheffé test for multiple comparisons was performed on the average miscounts to test for “flatness” by identifying homogeneous subsets in the different numerosity conditions. This comparison identifies the groups with similar performance (i.e., the group means were not statistically different). The results from this test indicate the following homogeneous subsets: for 50-ms displays, the numerosity range of 2-5 (average miscounts 0.02 to 0.07, \(p = 0.97\)); for 200-ms displays, the range of 2-6 (average...
miscounts 0.01 to 0.11, \( p = 0.07 \)); and for 350-ms displays, the range of 2-6 (average
miscounts 0.01 to 0.08, \( p = 0.27 \)).

These subsets were also identified using a trend analysis (for each duration
separately) that tested for significant deviations from linearity in the following
numerosity ranges: 2-4, 2-5, 2-6, 2-7, 2-8, and 2-9. When examining the average number
of miscounts, the last group that did not deviate from linearity was the 2-6 numerosity
range for both 50-ms and 200-ms displays. The last group for the 350-ms displays that
did not deviate from linearity was the 2-7 range. The breakpoint is taken to be after
numerosity 6 in 50-ms and 200-ms displays, and after 7 in the 350-ms displays. These
ranges can be considered the subitizing range when using this method of analysis. The
slope of average miscounts (Figure 3) in these subsets were relatively flat in the smaller
range and increased in the larger range. In the 50-ms displays, the slope for the 2-6
numerosity had an unstandardized\(^6\) \( \beta = 0.04 \) (\( p < 0.01 \)) and the 6-9 range had a slope of
\( \beta = 0.45 \) (\( p < 0.01 \)). In the 200-ms displays, the 2-6 range had a shallow slope as well,
\( \beta = 0.03 \) (\( p < 0.01 \)) and increased to \( \beta = 0.28 \) (\( p < 0.01 \)) in the 7-9 range. For the 350-ms
displays, the error slope in the 2-7 range was \( \beta = 0.03 \) (\( p < 0.01 \)) and increased in the 7-9
range to \( \beta = 0.30 \) (\( p < 0.01 \)).

The Bayesian test for flatness, however, identified a smaller subitizing range of 2-
4 for all display durations. This range was determined by identifying the largest
numerosity condition where the odds against the null hypothesis never reached 3:2. The
prior likelihood function for this numerosity indicates how vague the prior must be to
achieve this acceptance of the null (with the most vague prior being 0.25 and the most

\(^6\) The unstandardized \( \beta \) is reported here in order to correspond to the values on the charts, so that a
slope of 0.10, for example, indicates an average increase in counting error by 0.1 items on the
measure of “average miscounts”.
specific being 0.99). This value was: \( P = 0.95 \) for 50-ms displays, \( P = 0.50 \) for 200-ms displays, and \( P = 0.90 \) for 350-ms displays. This indicates a stronger case for flatness in the 2-4 numerosities for 50-ms and 350-ms displays than the 200-ms displays, since the prior probability must be “vaguer” to reject the null on 200-ms displays.

Finally, a simple linear function was computed to compare the numerosity response to the display numerosity. For small sets of 2-4 items and 2-6 items, the degree of linear fit between display numerosity and the observers’ responses was relatively high. See Table 1 for a summary of these results, including breakdowns by duration. This table shows results for the “typical subitizing range”, which refers to the range of 1-4 items found in previous experiments on subitizing (e.g., Trick & Pylyshyn, 1994b). The “enhanced subitizing range” refers to the subitizing range of 1-6 items that was observed in this “enumerating-by-pointing” task. There is little difference in the standardized-\( \beta \) between the ranges of 2-4 and 2-6, indicating similar performance in these ranges.

Localization accuracy

Error in localization accuracy is reported as the Euclidean distance between the center of a stimulus disc and a paired response marker. ANOVA results for the magnitude of location errors among the conditions indicate main effects for display duration \( (F(2,5232) = 35.8, \ p < 0.001, \ \eta^2_p = 0.602) \) and numerosity \( (F(6,5232) = 90.7, \ p < 0.001, \ \eta^2_p = 0.793) \), with a small interaction \( (F(12,5232) = 2.0, \ p < 0.05, \ \eta^2_p = 0.078) \). Figure 4 shows the average error distance in pixels and degrees of visual angle. Errors increased for larger numerosities and in the shortest presentation duration. A Bonferroni test for multiple comparisons indicate that performance in the 50-ms duration was significantly worse than the 200-ms duration for numerosities 2-5 and it was
significantly worse than the 350-ms duration for numerosities 2-7. There were no significant differences between the 200-ms and 350-ms conditions.

A regression analysis on the combined durations showed a larger increase (slope) of location errors with numerosity for displays with 2-6 items (unstandardized $\beta = 6.96$, $p < 0.001$) than for displays with 7-9 items ($\beta = 0.66$, $p = 0.57$). This difference in slope indicates the degree to which the rate in localization errors increased with numerosity, which was higher in the smaller numerosity range. This increase, however, may be a function of response order; that is, with each response, the localization errors grew due to additive effects of error, memory decay, or other sources. Localization errors will be examined in detail in Section 4.10, including an analysis of these errors as a function of response number (e.g., results from analyzing the location errors on the first-click response indicate that localizations errors do in fact increase as a function of increasing display numerosity and shorter presentation durations).

<table>
<thead>
<tr>
<th>Type of Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical subitizing range</td>
</tr>
<tr>
<td><strong>2-4 items</strong></td>
</tr>
<tr>
<td>50-ms</td>
</tr>
<tr>
<td>200-ms</td>
</tr>
<tr>
<td>350-ms</td>
</tr>
<tr>
<td>Combined</td>
</tr>
</tbody>
</table>

Table 1. Regression values when analyzing the correlation between the reported numerosity and the actual display numerosity in Experiment 1. Note: All values reported are standardized $\beta$ and adjusted $r^2$; all significance levels were $p < 0.001$. 
Figure 2. Proportion of trials enumerated correctly in Experiment 1.  
NOTE: All error bars in the report represent 95% C.I.

Figure 3. Average miscounts (absolute value) in Experiment 1.
Figure 4. Average distance between stimulus-response pairs in Experiment 1, in pixels (left y-axis) and visual angle in degrees (right y-axis).

Discussion

Experiment 1 indicates that increasing the duration of the stimulus presentation allows more accurate encoding of locations, with a marked advantage in the 200- and 350-ms displays over the 50-ms displays. This advantage was primarily observed for displays with less than seven discs. Enumeration performance for this range was not affected by the presentation duration; only displays with six or more items were affected by duration. Since pointing to discrete locations is required in order to correctly indicate the right number of items on the screen, the results indicate that individuating items is much faster than localizing them. This is consistent with the proposal that the visual indexing mechanism may be involved in this task, since items can be indexed before feature information, including location, is encoded. The results from this experiment also suggest that a rough spatial representation can be updated over time to produce more
accurate localization. (Note: See Section 4.10 for more in-depth analyses of the localization data, including errors as a function of response order.) Additionally, the enumeration performance observed in this experiment suggests a higher subitizing limit—up to six items—than traditional studies examining this phenomenon. The following experiments address this surprising finding of an increased subitizing range by directly comparing performance in the pointing task to other reporting methods.

4.3. Experiment 2A – Enumerating-by-Pointing vs. Cardinal Numeral Selection

The results of Experiment 1 suggest that the “pointing method” allows more items to be processed (and indirectly enumerated) in rapid enumeration than typically reported (e.g., Trick & Pylyshyn, 1994b). Since subitizing experiments have used a variety of methods, there is the possibility that the subitizing range increase found in this experiment may be due to aspects of the methodology other than pointing to recalled object locations. For example, it might be due to the luminance and timing of the stimulus and mask or the use of the mouse to indicate cardinality. Therefore, Experiment 2A compared the indirect “pointing method” of inferring how many items had been processed with a 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 Experiment 1. Since Experiment 2A involves the use of a different set of observers, the pointing method also was administered to the new subject population in order to provide a within-subjects comparison of the pointing response versus the symbolic numeral response (as well as replicating the results of Experiment 1). These results also were reported in Haladjian & Pylyshyn (2011).
Method

Experiment 2A consisted of two blocks. In the first (control) block, observers simply reported the number of objects by clicking on the corresponding Arabic numeral on a response screen that appeared 85-ms after the offset of the mask (with a black screen in between mask and response screen). The numerals (1-12) appeared on the screen in the form of a clock-like ring with a radius of \(~3.8^\circ\) (140 pixels) and centered on the location of the fixation cross. They were presented in 20-point Helvetica font in a gray color that was lighter than the background gray (RGB values of [200, 200, 200]). The selection cursor always appeared at the location of the fixation cross and observers moved it to click on a number using the computer mouse. The left image in Figure 5 depicts this response screen. The equal distances between the location of the response pointer on the screen and numeral option allowed a comparison of RTs for this response (however, the task instructions emphasized accuracy over speed and did not control for direction of motion required for each integer, so the RT is only a rough indication).

The second block used the same pointing task described in Experiment 1 (without the 350-ms condition). This block order (a numeral selection control block followed by a location selection block) was maintained to discourage the use of location pointing strategies in the control block. We anticipated no practice effect on enumeration performance in this arrangement since there were none in Experiment 1 (i.e., no practice effects on enumeration or localization performance). Additionally, RTs were recorded in both blocks. Eight numerosities (2-9) and two display durations (50 and 200 ms) were tested, for a total of 16 test conditions that were administered 10 times in each block. Twenty-two Rutgers University undergraduates completed both blocks in one 50-minute
session for course credit or payment. (Two more students completed this experiment, but their data were discarded because one student received an unbalanced number of trials and another student had unusually low performance in the numeral selection block.)

Figure 5. Depictions of the response screens used in Experiment 2A (left image) and Experiment 2B (right image).

Results

Enumeration accuracy

ANOVA were conducted on the proportion of trials correctly enumerated. In the numeral selection (control) block, there were main effects for display duration ($F(1,2800) = 53.8, p < 0.001$, $\eta^2_p = 0.719$) and numerosity ($F(6,2800) = 73.2, p < 0.001$, $\eta^2_p = 0.777$), with a significant interaction ($F(6,2800) = 5.7, p < 0.001$, $\eta^2_p = 0.213$).

In the pointing task (the same task used in Experiment 1), there were effects for display duration ($F(1,2800) = 29.8, p < 0.001$, $\eta^2_p = 0.586$) and numerosity ($F(6,2800) = 82.5, p < 0.001$, $\eta^2_p = 0.797$), with an interaction ($F(6,2800) = 7.9, p < 0.001$, $\eta^2_p = 0.273$).

These results follow the same patterns as those in Experiment 1.

Again, the absolute values of the miscounts were analyzed (undercounts accounted for 83% of miscounts in the numeral selection task and 92% in the pointing task). The ANOVA results for the numeral selection task indicate main effects for both
display duration \((F(1,2798) = 49.0, p < 0.001, \eta^2_p = 0.700)\) and numerosity \((F(6,2798) = 61.1, p < 0.001, \eta^2_p = 0.744)\), with an interaction \((F(6,2798) = 6.5, p < 0.001, \eta^2_p = 0.236)\). Also in the pointing task, there were effects for display duration \((F(1,2797) = 36.1, p < 0.001, \eta^2_p = 0.631)\) and numerosity \((F(6,2797) = 58.5, p < 0.001, \eta^2_p = 0.736)\), with an interaction \((F(6,2797) = 11.5, p < 0.001, \eta^2_p = 0.353)\). The results from the pointing task are similar to the results in Experiment 1.

Display duration had more of an effect on the numeral selection task than the pointing task. The Bonferroni test confirmed that the counting performance in the numeral selection task was significantly worse in the 50-ms condition than the 200-ms condition for displays with 3 items (MD = 0.05, \(p < 0.05\)), 5 items (MD = 0.17, \(p < 0.01\)), 6 items (MD = 0.13, \(p < 0.05\)), 7 items (MD = 0.17, \(p < 0.01\)), 8 items (MD = 0.31, \(p < 0.01\)), and 9 items (MD = 0.53, \(p < 0.01\)). In the pointing task, the differences between the 50-ms and 200-ms conditions were only significant for displays with 7 items (MD = 0.17, \(p < 0.01\)), 8 items (MD = 0.26, \(p < 0.01\)), and 9 items (MD = 0.45, \(p < 0.01\)). These results were comparable in the proportion of correct trials measure.

**Enumeration accuracy: Differences in reporting method**

To compare performance between the two reporting methodologies, ANOVAs were performed for each display duration separately. In these 2 × 7 ANOVAs, the independent factors were reporting method (explicit numerical response versus pointing response) and numerosity. In the 50-ms displays, ANOVA on the proportion of correct trials showed main effects for reporting method \((F(1,2800) = 19.1, p < 0.001, \eta^2_p = 0.476)\) and numerosity \((F(6,2800) = 108.5, p < 0.001, \eta^2_p = 0.838)\), with an interaction \((F(6,2800) = 7.8, p < 0.001, \eta^2_p = 0.271)\). In the 200-ms displays, there were
also main effects for reporting method \((F(1,2800) = 8.0, p < 0.01, \eta^2_p = 0.276)\) and numerosity \((F(6,2800) = 56.9, p < 0.001, \eta^2_p = 0.730)\), with an interaction \((F(6,2800) = 5.8, p < 0.001, \eta^2_p = 0.215)\). See Figure 6.

ANOVA results for the average miscounts (absolute value) also indicated main effects for reporting method \((F(1,2795) = 19.4, p < 0.001, \eta^2_p = 0.480)\) and numerosity \((F(6,2795) = 76.3, p < 0.001, \eta^2_p = 0.784)\), with an interaction \((F(6,2795) = 6.4, p < 0.001, \eta^2_p = 0.234)\). In the 200-ms displays, there were also main effects for reporting method \((F(1,2800) = 7.9, p < 0.01, \eta^2_p = 0.273)\) and numerosity \((F(6,2800) = 50.7, p < 0.001, \eta^2_p = 0.707)\), with an interaction \((F(6,2800) = 5.3, p < 0.001, \eta^2_p = 0.200)\). See Figure 7. Bonferroni comparisons between reporting methods showed a significant advantage for the pointing task (over the numeral selection task) in the 50-ms condition for displays with 5 items \((\text{MD} = 0.16, p < 0.05)\), 6 items \((\text{MD} = 0.33, p < 0.01)\), and 7 items \((\text{MD} = 0.14, p < 0.01)\). For the 200-ms displays, the advantage for the pointing task was only evident in displays with 6 items \((\text{MD} = 0.21, p < 0.01)\) and 7 items \((\text{MD} = 0.13, p < 0.01)\).

**Subitizing range**

The Scheffé test was used to identify subsets of performance measures that were homogeneous (or “flat”) in their overall magnitudes. The results from this test indicate the following homogeneous subsets in the numeral selection task (control): for 50-ms displays, the numerosity range of 2-4 (average miscounts 0.02 to 0.06, \(p = 0.99)\); and for 200-ms displays, the range of 2-5 (average miscounts 0.01 to 0.05, \(p = 0.86)\). The following subsets were identified in the pointing task: for 50-ms displays, the range of 2-6 (average miscounts 0.02 to 0.07, \(p = 0.89)\); and for 200-ms displays, the range of 2-6...
(average miscounts 0.01 to 0.06, \( p = 0.65 \)). These subsets were the same for proportion of correct trials. These ranges were confirmed in a trend analysis by testing for significant deviations from linearity for each duration and reporting method separately (for these numerosity ranges: 2-4, 2-5, 2-6, 2-7, 2-8, 2-9). In the control task, the largest range that did not deviate from linearity was the 2-4 range in 50-ms displays (unstandardized \( \beta = 0.02, \text{N.S.} \)) and the 2-5 range in 200-ms displays (\( \beta = 0.02, p < 0.01 \)). In the pointing task, the largest range that did not deviate from linearity was the 2-6 range in both 50-ms (\( \beta = 0.01, p < 0.01 \)) and 200-ms displays (\( \beta = 0.01, p < 0.01 \)). The Bayesian test for flatness identified in the control task a subitizing range of 2-4 for both 50-ms displays (\( P = 0.90 \)) and 200-ms displays (\( P = 0.95 \)). In the pointing task, this range was 2-5 for both 50-ms displays (\( P = 0.85 \)) and 200-ms displays (\( P = 0.90 \)).

Finally, a simple linear function was computed to compare the numerosity response to the display numerosity. When computed for the “enhanced subitizing range”, the degree of linear fit between display numerosity and numerical responses were higher for the pointing task than the numeral selection task. See Table 2 for a summary of these results.

Response time results

The RT for making the numeral responses in the first block were examined. Outlier durations of >5 seconds were excluded from the analysis (mean RT = 1.416, SD = 0.866). ANOVA results on the RTs in the numeral selection (control) block revealed a main effect for display duration (\( F(1,2760) = 43.0, p < 0.001, \eta^2_p = 0.668 \)) and numerosity (\( F(6,2760) = 116.7, p < 0.001, \eta^2_p = 0.847 \)), with no interaction (\( F(6,2760) = 0.6, p = 0.70, \eta^2_p = 0.029 \)). A trend analysis on the RT data indicates that
all of the testable ranges significantly deviated from linearity; therefore, it is likely that
the 2-3 may be linear for both 50-ms displays ($\beta = 0.02, \text{N.S.}$) and 200-ms displays
($\beta = 0.01, \text{N.S.}$). These results are plotted in Figure 8. The RT for the “first click” in the
pointing task is also plotted on these charts, although they show no interesting trends:
there was no effect of numerosity and RTs averaged to be the same across all
numerosities, as did the response intervals between each click.

Localization accuracy

Since localization errors will be examined in detail in Section 4.10, only the
overall ANOVA results will be reported here (also true for subsequent experiments).
ANOVA results for the magnitude of location errors were similar to those of Experiment
1 and indicate main effects for display duration ($F(1,2549) = 8.9, p < 0.01, \eta^2_p = 0.283$)
and numerosity ($F(6,2549) = 28.0, p < 0.001, \eta^2_p = 0.554$), but with no interaction
($F(6,2549) = 1.3, p = 0.26, \eta^2_p = 0.054$). Location errors as a function of numerosity are
plotted in Figure 9.

<table>
<thead>
<tr>
<th>Task</th>
<th>Typical subitizing range</th>
<th>Enhanced subitizing range</th>
<th>Full range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-4 items</td>
<td>5-9 items</td>
<td>2-6 items</td>
</tr>
<tr>
<td>Pointing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>$\beta = 0.976, r^2 = 0.953, \beta = 0.857, r^2 = 0.734$</td>
<td>$\beta = 0.986, r^2 = 0.973, \beta = 0.528, r^2 = 0.278$</td>
<td>$\beta = 0.966, r^2 = 0.933$</td>
</tr>
<tr>
<td>200-ms</td>
<td>$\beta = 0.992, r^2 = 0.984, \beta = 0.911, r^2 = 0.829$</td>
<td>$\beta = 0.993, r^2 = 0.987, \beta = 0.664, r^2 = 0.440$</td>
<td>$\beta = 0.979, r^2 = 0.958$</td>
</tr>
<tr>
<td>Numeral</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>$\beta = 0.949, r^2 = 0.901, \beta = 0.807, r^2 = 0.651$</td>
<td>$\beta = 0.943, r^2 = 0.890, \beta = 0.526, r^2 = 0.275$</td>
<td>$\beta = 0.948, r^2 = 0.899$</td>
</tr>
<tr>
<td>200-ms</td>
<td>$\beta = 0.992, r^2 = 0.985, \beta = 0.880, r^2 = 0.774$</td>
<td>$\beta = 0.979, r^2 = 0.959, \beta = 0.649, r^2 = 0.420$</td>
<td>$\beta = 0.970, r^2 = 0.942$</td>
</tr>
</tbody>
</table>

Table 2. Regression values when analyzing the correlation between the reported numerosity and
the actual display numerosity in Experiment 2A. Note: All values reported are standardized $\beta$ and
adjusted $r^2$; all significance levels were $p < 0.001$. 
Figure 6. Proportion of trials enumerated correctly for pointing and numeral report conditions in Experiment 2A (50-ms on left and 200-ms on right).

Figure 7. Average miscounts for pointing and numeral report conditions in Experiment 2A (50-ms on left and 200-ms on right).

Figure 8. Response times; for the pointing task, RT is reported for the first click only in Experiment 2A (50-ms on left and 200-ms on right).
Discussion

The results from Experiment 2A suggest 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 cardinality of the set of dots. Additionally, RTs from the numeral selection task showed typical subitizing effects, where the RT slope was small in the subitizing range of 2-3 items and increased substantially for larger sets. Thus, both the RT and enumeration results indicate that subitizing performance in the numeral selection task was similar to that found in previous studies. The constant RT for the first click response in the pointing task suggests that the observers were not computing the display numerosity prior to making their response (if they did, the RT should increase with numerosity). Enumeration accuracy was higher in the pointing task for 50-ms displays with numerosities of 5-7, and for the 200-ms displays with
numerosities 6-7. The additional encoding period in the 200-ms displays seems to improve enumeration accuracy for displays with 5 items, but there was a marked advantage for the location-pointing method for the range of 5-8 items in the 50-ms condition and 6-7 items in the 200-ms condition. This indicates a small overall advantage for reporting numerosity nonverbally. Although this trend cannot be confirmed reliably for individual observers (since only 10 trials were collected per condition per participant), this trend does seem to hold for a majority of the observers (see Figure A in the Appendix for charts displaying overall performance for individual subjects).

One possible criticism with the design of the response screen is the location of where the numerals for selection appeared. Since they appeared in a ring around the central fixation, their appearance may have interfered with observers’ representation of the stimulus and thus reduced performance. In order to test this possibility, Experiment 2B was devised in a way that avoids this possible interference by presenting the numerals for response selection at the bottom of the screen where no test stimuli appeared.

4.4. Experiment 2B - Enumerating-by-Pointing vs. Row Cardinal Selection

Experiment 2B is a replication of Experiment 2A except an alternate response screen was utilized in the control condition (see the image on the right in Figure 5). This experiment presented the selection numerals in a single row at the bottom of the screen. The goal of this presentation was to ensure that the results from Experiment 2A were due to the response method (i.e., selecting a cardinal numeral instead of pointing to locations) and not because the central response screen interfered with the stimulus representation and thus produced more recall errors.
Method

The design of Experiment 2B is identical to Experiment 2A, except that in the control block the Arabic numerals (from 1-10) were presented in a single row that appeared 150 pixels (~4°) above the bottom edge of the screen (or 362 pixels (~10°) below the central fixation) in 20-point Helvetica font. This response screen appeared 85-ms after the offset of the mask (with a black screen in between). This allowed us to compare performance in the numeral selection block to the location-pointing block in a version where the response screen did not occupy the same region as the targets on the stimuli displays. Twenty Rutgers University undergraduates completed both blocks in one 50-minute session for course credit or payment. (Note: Three subjects were removed because of extremely low or variable performance in the numeral selection block, which exaggerated the overall differences between the two conditions.)

Results

Enumeration accuracy

ANOVA were conducted on the proportion of trials with correct enumeration. In the control block, enumeration accuracy as proportion of trials correctly enumerated showed main effects for both display duration ($F(1,2520) = 64.2, p < 0.001$, $\eta^2_p = 0.772$) and numerosity ($F(6,2520) = 47.4, p < 0.001, \eta^2_p = 0.714$), with an interaction ($F(6,2520) = 7.8, p < 0.001, \eta^2_p = 0.292$). For the pointing task, there were effects for display duration ($F(1,2520) = 25.0, p < 0.001, \eta^2_p = 0.568$) and numerosity ($F(6,2520) = 48.2, p < 0.001, \eta^2_p = 0.717$), with an interaction ($F(6,2520) = 7.8, p < 0.001, \eta^2_p = 0.291$). See Figure B in the Appendix for charts displaying overall performance for individual subjects.
The absolute values of the miscounts were analyzed (84% of miscounts in the numeral selection task were undercounts, and 92% of errors in the pointing task were undercounts). ANOVA on the average miscounts in the control condition showed main effects for display duration ($F(1,2519) = 51.2$, $p < 0.001$, $\eta^2_p = 0.729$) and numerosity ($F(6,2519) = 37.3$, $p < 0.001$, $\eta^2_p = 0.664$), with an interaction ($F(6,2519) = 11.7$, $p < 0.001$, $\eta^2_p = 0.382$). Results from the pointing task also showed effects for display duration ($F(1,2520) = 23.6$, $p < 0.001$, $\eta^2_p = 0.554$) and numerosity ($F(6,2520) = 40.1$, $p < 0.001$, $\eta^2_p = 0.678$), with an interaction ($F(6,2520) = 9.7$, $p < 0.001$, $\eta^2_p = 0.339$). These results were similar to Experiment 2A.

Again, display duration had more of an effect on the numeral selection task than the pointing task. A Bonferroni test confirmed that the counting performance in the numeral selection task was significantly worse in the 50-ms condition than the 200-ms condition for displays with 2 items (MD = 0.04, $p < 0.05$), 3 items (MD = 0.04, $p < 0.05$), 4 items (MD = 0.05, $p < 0.01$), 6 items (MD = 0.14, $p < 0.01$), 7 items (MD = 0.20, $p < 0.01$), 8 items (MD = 0.47, $p < 0.01$), and 9 items (MD = 0.57, $p < 0.01$). In the pointing task, the differences between the 50-ms and 200-ms conditions were only significant for displays with 6 items (MD = 0.13, $p < 0.01$), 7 items (MD = 0.19, $p < 0.01$), 8 items (MD = 0.24, $p < 0.01$), and 9 items (MD = 0.39, $p < 0.01$). These results were comparable in the proportion of correct trials measure.

**Enumeration accuracy: Differences in reporting method**

To examine how reporting method affected enumeration performance, ANOVAs were performed with reporting method and numerosity as factors for each display duration. In the 50-ms condition, ANOVA on the proportion of correct trials showed
main effects for reporting method ($F(1,2520) = 6.0, p < 0.05, \eta_P^2 = 0.241$) and numerosity ($F(6,2520) = 68.2, p < 0.001, \eta_P^2 = 0.782$), with an interaction ($F(6,2520) = 2.2, p < 0.05, \eta_P^2 = 0.104$). In the 200-ms condition, there was no effect of reporting method ($F(1,2520) = 0.7, p = 0.41, \eta_P^2 = 0.036$), but there was an effect for numerosity ($F(6,2520) = 39.0, p < 0.001, \eta_P^2 = 0.672$), with an interaction ($F(6,2520) = 2.8, p < 0.05, \eta_P^2 = 0.130$). See Figure 10.

ANOVAs for the average miscounts (absolute value) in the 50-ms condition indicate main effects for reporting method ($F(1,2519) = 5.9, p < 0.05, \eta_P^2 = 0.236$) and numerosity ($F(6,2519) = 48.5, p < 0.001, \eta_P^2 = 0.719$), with no interaction ($F(6,2519) = 1.8, p = 0.10, \eta_P^2 = 0.087$). In the 200-ms displays, there was no effect of reporting method ($F(1,2520) = 0.6, p = 0.46, \eta_P^2 = 0.029$), but there was an effect for numerosity ($F(6,2520) = 36.5, p < 0.001, \eta_P^2 = 0.657$), with an interaction ($F(6,2520) = 2.8, p < 0.05, \eta_P^2 = 0.127$). See Figure 11. Since there was an effect for reporting method and an interaction, a Bonferroni test comparing miscounts between the two reporting methods was conducted. The results indicate a significant advantage for the pointing task (over the numeral selection task) in the 50-ms condition for displays with 5 items (MD = 0.12, $p < 0.01$), 6 items (MD = 0.14, $p < 0.01$), and 8 items (MD = 0.18, $p < 0.01$). For the 200-ms displays, the advantage for the pointing task was only evident in displays with 6 items (MD = 0.13, $p < 0.01$).

**Subitizing range**

The Scheffé test was conducted on the average miscounts to identify homogeneous subsets. The results from this test indicate the following homogeneous subsets in the numeral selection task (control): for 50-ms displays, the numerosity range
of 2-5 (average miscounts 0.04 to 0.13, $p = 0.74$); and for 200-ms displays, the range of 2-5 (average miscounts 0.01 to 0.06, $p = 0.60$). The following subsets were identified in the pointing task: for 50-ms displays, the range of 2-5 (average miscounts 0.01 to 0.02, $p = 0.99$); and for 200-ms displays, the range of 2-6 (average miscounts 0.01 to 0.04, $p = 0.95$). These subsets were the same for proportion of correct trials. These ranges were confirmed in a trend analysis by testing for significant deviations from linearity for each duration and reporting method separately among different numerosity ranges. In the control task, the largest range that did not significantly deviate from linearity was the 2-5 range for 50-ms displays (unstandardized $\beta = 0.03$, $p < 0.05$) and the 2-4 range for 200-ms displays ($\beta = 0.01$, N.S.). In the pointing task, these ranges were 2-5 for the 50-ms displays ($\beta = 0.0$, N.S.) and the 2-6 range for 200-ms displays ($\beta = 0.01$, $p < 0.01$). The result that the subitizing range in the control task appears to be worse for 200-ms displays than 50-ms displays may be due to variability in individual subject responses. The Bayesian test for flatness identified in the control task a subitizing range of 2-4 for both 50-ms displays ($P = 0.95$) and 200-ms displays ($P = 0.99$). In the pointing task, this range was 2-5 for both 50-ms displays ($P = 0.99$) and 200-ms displays ($P = 0.85$), although the 200-ms displays had a range of 2-6 for $P = 0.75$.

A simple linear function again was computed to examine the relationship between the numerosity response to the display numerosity. When computed for the “enhanced subitizing range”, the degree of linear fit between display numerosity and numerical responses were higher for the pointing task than the numeral selection task. See Table 3 for a summary of these results.
Response time results

The RTs for making the numeral responses in the first block were analyzed after excluding outliers > 5 seconds (mean RT = 1.636, SD = 0.942). ANOVA results in the numeral selection block indicate main effects for both display duration \( (F(1,2455) = 11.6, p < 0.01, \eta^2_p = 0.377) \) and numerosity \( (F(6,2455) = 127.3, p < 0.001, \eta^2_p = 0.870) \), with an interaction \( (F(6,2455) = 3.0, p < 0.01, \eta^2_p = 0.134) \). A trend analysis on the RT data indicates that the breakpoint from linearity occurred after the numerosity range of 2-4 \( (\beta = 0.10, p < 0.01) \) for 50-ms displays and 2-3 \( (\beta = -0.05, \text{N.S.}) \) for 200-ms displays.

Again, no interesting trends in RT were detected in the pointing task (for first click or time between clicks). See Figure 12. Note that typical subitizing slope is still observed in this experiment even though the distance to click on extreme values (1-3 or 8-10) was greater than that for middle values (4-7).

Localization accuracy

ANOVA results for location errors were similar to the previous results and indicate main effects for duration \( (F(1,2296) = 6.1, p < 0.05, \eta^2_p = 0.241) \) and numerosity \( (F(6,2296) = 28.9, p < 0.001, \eta^2_p = 0.594) \), with no interaction \( (F(6,2296) = 0.6, p = 0.76, \eta^2_p = 0.028) \). See Figure 13.

Differences between Experiments 2A & 2B

The numeral selection responses from Experiments 2A & 2B were tested for differences in performance by including them as factors in an ANOVA. Ideally, performance should be comparable in the two experiments. When examining the ANOVA results for the proportion of correct trials, there were no significant effects for Experiment in both the 50-ms condition \( (F(1,2940) = 3.3, p = 0.12, \eta^2_p = 0.348) \) and the
200-ms condition \((F(1,2940) = 4.6, p = 0.08, \eta^2_p = 0.433)\). The ANOVA results for the average miscounts also indicate no main effect for experiment type in both the 50-ms condition \((F(1,2937) = 4.2, p = 0.09, \eta^2_p = 0.414)\) and the 200-ms condition \((F(1,2940) = 4.5, p = 0.08, \eta^2_p = 0.426)\). These results were nearly significant, however, and this difference may be due to the shortened subitizing range in the 50-ms pointing task (where the range was up to 5 items instead of the 6-item limit in Experiment 2A).

<table>
<thead>
<tr>
<th>Task</th>
<th>Typical subitizing range</th>
<th>Enhanced subitizing range</th>
<th>Full range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-4 items</td>
<td>2-6 items</td>
<td>2-9 items</td>
</tr>
<tr>
<td></td>
<td>5-9 items</td>
<td>7-9 items</td>
<td></td>
</tr>
<tr>
<td><strong>Pointing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>(\beta = .989, r^2 = .979)</td>
<td>(\beta = .988, r^2 = .976)</td>
<td>(\beta = .967, r^2 = .936)</td>
</tr>
<tr>
<td>200-ms</td>
<td>(\beta = .994, r^2 = .988)</td>
<td>(\beta = .996, r^2 = .991)</td>
<td>(\beta = .979, r^2 = .959)</td>
</tr>
<tr>
<td><strong>Numeral</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>(\beta = .956, r^2 = .913)</td>
<td>(\beta = .801, r^2 = .642)</td>
<td>(\beta = .950, r^2 = .903)</td>
</tr>
<tr>
<td>200-ms</td>
<td>(\beta = .996, r^2 = .993)</td>
<td>(\beta = .918, r^2 = .842)</td>
<td>(\beta = .980, r^2 = .961)</td>
</tr>
</tbody>
</table>

Table 3. Regression values when analyzing the correlation between the reported numerosity and the actual display numerosity in Experiment 2B. Note: All values reported are standardized \(\beta\) and adjusted \(r^2\); all significance levels were \(p < 0.001\).

Figure 10. Proportion of trials enumerated correctly for pointing and numeral (row) report conditions in Experiment 2B (50-ms on left and 200-ms on right).
Figure 11. Average miscounts for pointing and numeral (row) report conditions in Experiment 2B (50-ms on left and 200-ms on right).

Figure 12. Response times; in the location-marking condition RT is for the first click in Experiment 2B (50-ms on left and 200-ms on right).
Discussion

Experiment 2B also showed an advantage for the nonverbal response method of pointing to locations over the numeral selection task. This advantage was not as strong in the 50-ms condition due to lower performance in the pointing task. The counting errors in Experiments 2A and 2B were compared and there was no difference in the overall performance between these two versions in the numeral selection task (or, less interestingly, the pointing task). The results from both of these experiments indicate that the difference in performance between the two reporting methodologies were not due to interference from the response display but rather to the method of responding. There is an advantage to pointing to locations instead of reporting a cardinal number regardless of where the numeral selection array was presented (i.e., center or bottom of the screen). The trend analysis identified a higher subitizing range for the 200-ms displays in the pointing task (2-6 range) than in the numeral selection task (2-4 range), but there was no
advantage for the 50-ms displays (2-5 range for both tasks). The reason for an advantage in the nonverbal reporting method of numerosity could be due to a variety of reasons, such as the pointing nature of the task engaging a motor memory to supplement capacity or from a change in the allocation of resources required when assigning a numeral label. The possibility that the nonverbal nature of the pointing task accounts for the performance difference is addressed in the subsequent experiments.

4.5. Experiment 3A: Enumerating-by-Pointing vs. Grid Tally

Since there seems to be an advantage for a nonverbal report of numerosity, Experiment 3A included an alternate method for nonverbally reporting numerosity. In the control block, observers simply clicked the mouse (no pointer was visible) to “tally” their response instead of reporting the numerosity by clicking on an Arabic numeral as in Experiments 2A or 2B. This allowed a direct within-subjects comparison of performance when observers reported numerosity by tallying the number of discs seen on the stimulus display and when they reported numerosity by marking disc locations. This aims to address the question of whether the enhanced subitizing range is due to the non-symbolic nature of the response or if it also requires a spatial correspondence.

Method

Experiment 3A was identical in design to those in Experiment 2 except for the method of reporting numerosity in the first block. In this control block, observers simply clicked a mouse to “tally” their response (no pointer was visible). Each click made one black “X” appear in a 3x3 borderless grid at the center of the screen. This version of the control provided observers with a visual representation of numerosity during the response
stage but in a non-cardinal and spatially-irrelevant way. The standard pointing task was administered in the second block. Experiment 3A aimed to test whether clicking to spatial locations aided performance or if the act of “clicking” that produced a spatially-irrelevant record of each click would yield the same advantage. Twenty-one Rutgers students completed both blocks in a 50-minute session for course credit.

Results

Enumeration accuracy

ANOVA was conducted on the proportion of trials enumerated correctly. In the control block where observers responded by tallying their response, there were main effects for duration \((F(1,2702) = 23.5, p < 0.001, \eta^2_p = 0.539)\) and numerosity \((F(6,2702) = 62.2, p < 0.001, \eta^2_p = 0.756)\), with an interaction \((F(6,2702) = 8.7, p < 0.001, \eta^2_p = 0.302)\). For the pointing task, there also were effects for duration \((F(1,2702) = 12.5, p < 0.01, \eta^2_p = 0.383)\) and numerosity \((F(6,2702) = 90.7, p < 0.001, \eta^2_p = 0.819)\), with an interaction \((F(6,2702) = 6.7, p < 0.001, \eta^2_p = 0.249)\). See Figure C in the Appendix for charts displaying overall performance for individual subjects.

The absolute values of the miscounts were analyzed (80% of miscounts in the control task were undercounts, and 88% of errors in the pointing task were undercounts). ANOVA results from the control block showed effects for display duration \((F(1,2701) = 19.7, p < 0.001, \eta^2_p = 0.496)\) and numerosity \((F(6,2701) = 52.5, p < 0.001, \eta^2_p = 0.723)\), with an interaction \((F(6,2701) = 8.1, p < 0.001, \eta^2_p = 0.286)\). The pointing task also showed effects for display duration \((F(1,2701) = 16.3, p < 0.001, \eta^2_p = 0.449)\) and numerosity \((F(6,2701) = 63.9, p < 0.001, \eta^2_p = 0.761)\), with an interaction \((F(6,2701) = 8.2, p < 0.001, \eta^2_p = 0.291)\).
Display duration had more of an effect on the control (tally) task than the pointing task. The Bonferroni test indicates that the counting performance in the control (tally) task was significantly worse in the 50-ms condition than the 200-ms condition for displays with 6 items (MD = 0.07, p < 0.05), 7 items (MD = 0.21, p < 0.01), 8 items (MD = 0.21, p < 0.01), and 9 items (MD = 0.25, p < 0.01). In the pointing task, the differences between the 50-ms and 200-ms conditions were only significant for displays with 8 items (MD = 0.26, p < 0.01) and 9 items (MD = 0.33, p < 0.01). These results were comparable in the proportion of correct trials measure.

*Enumeration accuracy: Differences in reporting method*

To examine how reporting method affected enumeration performance, ANOVAs were performed for the two display durations separately with reporting method and numerosity as factors. In the 50-ms displays, ANOVA on the proportion of correct trials showed no effect for reporting method ($F(1,2702) = 2.8, p = 0.11, \eta^2_p = 0.121$), but there was an effect of numerosity ($F(6,2702) = 92.4, p < 0.001, \eta^2_p = 0.822$), with an interaction ($F(6,2702) = 6.6, p < 0.001, \eta^2_p = 0.247$). In the 200-ms displays, there was no effect of reporting method ($F(1,2702) = 0.5, p = 0.46, \eta^2_p = 0.026$), but there was a main effect for numerosity ($F(6,2702) = 57.8, p < 0.001, \eta^2_p = 0.739$), with no interaction ($F(6,2702) = 1.9, p = 0.09, \eta^2_p = 0.085$). See Figure 14.

ANOVA for the average miscounts in the 50-ms condition also revealed no main effects for reporting method ($F(1,2701) = 2.5, p = 0.13, \eta^2_p = 0.110$) but there was an effect of numerosity ($F(6,2701) = 72.4, p < 0.001, \eta^2_p = 0.783$), with an interaction ($F(6,2701) = 7.3, p < 0.001, \eta^2_p = 0.265$). In the 200-ms displays, there was no effect of reporting method ($F(1,2701) = 0.2, p = 0.69, \eta^2_p = 0.008$), but there was a main effect
for numerosity \((F(6,2701) = 44.0, p < 0.001, \eta^2_p = 0.687)\), with no interaction \((F(6,2701) = 1.7, p = 0.14, \eta^2_p = 0.076)\). See Figure 15. Since there was an interaction between reporting method and numerosity, Bonferroni comparisons of counting errors between reporting methods were conducted. Results indicate a significant advantage for the pointing task (over the control tally task) in the 50-ms condition for displays with 6 items \((MD = 0.09, p < 0.01)\) and 7 items \((MD = 0.18, p < 0.01)\). There were no differences in the 200-ms conditions.

**Subitizing range**

The Scheffé test for multiple comparisons was conducted on the average miscounts in order to identify homogeneous subsets. The results from this test indicate the following homogeneous subsets in the tally task (control): for 50-ms displays, the numerosity range of 2-5 (average miscounts 0.02 to 0.04, \(p = 0.99\)); and for 200-ms displays, the range of 2-5 (average miscounts 0.01 to 0.03, \(p = 0.99\)). The following subsets were identified in the pointing task: for 50-ms displays, the range of 2-6 (average miscounts 0.01 to 0.09, \(p = 0.14\)); and for 200-ms displays, the range of 2-6 (average miscounts 0.01 to 0.07, \(p = 0.34\)). These subsets were the same for proportion of correct trials. These ranges were confirmed in a trend analysis by testing for significant deviations from linearity for each duration and reporting method separately in various numerosity ranges. In the control task, the largest range whose trend did not significantly deviate from linearity was the 2-5 range for both 50-ms displays \((\beta = 0.01, \text{N.S.})\) and 200-ms displays \((\beta = 0.01, \text{N.S.})\). In the pointing task, the last range without deviations from linearity were the 2-6 range for both 50-ms \((\beta = 0.02, p < 0.01)\) and 200-ms displays \((\beta = 0.02, p < 0.01)\). The Bayesian test for flatness identified in the control task a
subitizing range of 2-5 for 50-ms displays ($P = 0.975$) and 2-4 for 200-ms displays ($P = 0.95$). In the pointing task, this range was 2-5 for 50-ms displays ($P = 0.75$) and 2-4 for 200-ms displays ($P = 0.99$).

Finally, the simple linear function comparing the numerosity response to the stimulus numerosity indicates similar performance in the “typical” and “enhanced” subitizing ranges, with a slightly better linear fit in the pointing task. See Table 4 for a summary of these results.

Localization accuracy

ANOVA results for the magnitude of location errors indicate no main effects for display duration ($F(1,2315) = 0.1, p = 0.93, \text{eta}_p^2 = 0.001$), but there were significant main effects for numerosity ($F(6,2315) = 31.4, p < 0.001, \text{eta}_p^2 = 0.592$), with interactions ($F(6,2315) = 2.2, p < 0.05, \text{eta}_p^2 = 0.092$). See Figure 16.

<table>
<thead>
<tr>
<th>Task</th>
<th>Typical subitizing range</th>
<th>Enhanced subitizing range</th>
<th>Full range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-4 items</td>
<td>5-9 items</td>
<td>2-6 items</td>
</tr>
<tr>
<td>Pointing 50-ms</td>
<td>$\beta = .989, r^2 = .978$</td>
<td>$\beta = .888, r^2 = .788$</td>
<td>$\beta = .991, r^2 = .982$</td>
</tr>
<tr>
<td>200-ms</td>
<td>$\beta = .994, r^2 = .989$</td>
<td>$\beta = .922, r^2 = .850$</td>
<td>$\beta = .993, r^2 = .986$</td>
</tr>
<tr>
<td>Tally 50-ms</td>
<td>$\beta = .962, r^2 = .926$</td>
<td>$\beta = .877, r^2 = .769$</td>
<td>$\beta = .981, r^2 = .962$</td>
</tr>
<tr>
<td>200-ms</td>
<td>$\beta = .986, r^2 = .971$</td>
<td>$\beta = .905, r^2 = .819$</td>
<td>$\beta = .989, r^2 = .979$</td>
</tr>
</tbody>
</table>

Table 4. Regression values when analyzing the correlation between the reported numerosity and the actual display numerosity in Experiment 3A. Note: All values reported are standardized $\beta$ and adjusted $r^2$; all significance levels were $p < 0.001$. 
Figure 14. Proportion of trials enumerated correctly for pointing and tally (grid) report conditions in Experiment 3A (50-ms on left and 200-ms on right).

Figure 15. Average miscounts for pointing and tally (grid) report conditions in Experiment 3A (50-ms on left and 200-ms on right)
Discussion

There is a small effect of reporting method within Experiment 3A, suggesting that enumeration performance when reporting numerosity by pointing to locations was slightly better than keeping a spatially-irrelevant tally, but only in the 50-ms display condition for numerosities of 6 and 7 items (confirmed by pairwise comparisons). The trend analyses identified a higher subitizing range in the pointing task (2-6 items) than the tally task (2-5 items) for both durations. Although there is slightly better performance in the pointing task, the tally method still appears to be better than reporting by selecting a cardinal symbol for numerosity. An external “memory” served by the visual tally of already-counted items in the observers’ representations may be helping the enumeration process. These results also indicate that the 200-ms displays produce more stable representations for enumeration, which may be due to the limited encoding time in the 50-ms display presentations. Similar to the motivation of Experiment 2B, Experiment 3B
was designed to provide a tally but at the bottom of the screen instead of the center of the screen to reduce possible interference and confirm the small advantage for pointing to locations was not due to the response display position in the tally task.

4.6. Experiment 3B: Enumerating-by-Pointing vs. Row Tally

Experiment 3B is a replication of Experiment 3A except an alternate response screen was used in the control block. Here, the “tally” response appeared in a single row at the bottom of the screen to ensure that the central tally in Experiment 3A did not interfere with the response stage.

Method

Experiment 3B was identical to Experiment 3A except for the configuration of the tally in the control block. When observers clicked on the screen to “tally” their response, each click produced one black “X” marker in a single row at the bottom of the screen. This tally appeared 150 pixels (~4°) above the bottom edge of the screen and 362 pixels (~10°) below the central fixation—a location where stimulus discs would not appear (similar to Experiment 2B). Twenty-one Rutgers undergraduates completed both blocks in one 50-minute session for course credit.

Results

Enumeration accuracy

ANOVA were conducted on the proportion of trials with correct enumeration. In the control block with the row tally, there were effects for both display duration \( (F(1,2646) = 20.7, p < 0.001, \eta^2_p = 0.509) \) and numerosity \( (F(6,2646) = 40.8, p < 0.001, \eta^2_p = 0.671) \), with an interaction \( (F(6,2646) = 7.5 p < 0.001, \eta^2_p = 0.272) \). For the
pointing task, there were effects for display duration \( (F(1,2646) = 13.5, p < 0.001, \eta^2_p = 0.402) \) and numerosity \( (F(6,2646) = 64.0, p < 0.001, \eta^2_p = 0.762) \), with an interaction \( (F(6,2646) = 4.8, p < 0.001, \eta^2_p = 0.194) \). See Figure D in the Appendix for charts displaying overall performance for individual subjects.

The absolute values of the miscounts were analyzed (79% of miscounts in the control task were undercounts, and 81% of errors in the pointing task were undercounts). ANOVA on the average miscounts showed effects for display duration \( (F(1,2644) = 17.2, p < 0.001, \eta^2_p = 0.463) \) and numerosity \( (F(6,2644) = 35.4, p < 0.001, \eta^2_p = 0.639) \), with an interaction \( (F(6,2644) = 11.2, p < 0.001, \eta^2_p = 0.358) \). For the pointing task, there were effects for display duration \( (F(1,2646) = 13.8, p < 0.001, \eta^2_p = 0.409) \) and numerosity \( (F(6,2646) = 50.6, p < 0.001, \eta^2_p = 0.717) \), with an interaction \( (F(6,2646) = 6.6, p < 0.001, \eta^2_p = 0.249) \).

A Bonferroni test on average miscounts indicates that performance in the control (tally) task was significantly worse in the 50-ms condition than the 200-ms condition for displays with 6 items (MD = 0.09, \( p < 0.05 \)), 7 items (MD = 0.14, \( p < 0.01 \)), 8 items (MD = 0.33, \( p < 0.01 \)), and 9 items (MD = 0.37, \( p < 0.01 \)). In the pointing task, the differences between the 50-ms and 200-ms conditions were significant for displays with 7 items (MD = 0.17, \( p < 0.01 \)), 8 items (MD = 0.24, \( p < 0.01 \)), and 9 items (MD = 0.29, \( p < 0.01 \)). These results were comparable for the proportion of correct trials, but there was an additional difference in the control task for 4-item displays (MD = 0.04, \( p < 0.05 \)).

**Enumeration accuracy: Differences in reporting method**

To examine how reporting method affected enumeration performance, ANOVAs were performed for the two display durations separately with reporting method and
numerosity as factors. In the 50-ms displays, ANOVA on the proportion of correct trials showed no effect of reporting method ($F(1, 2646) = 0.2, p = 0.66, \eta^2_p = 0.010$), but there was an effect of numerosity ($F(6, 2646) = 71.7, p < 0.001, \eta^2_p = 0.782$), without an interaction ($F(6, 2646) = 0.7, p = 0.68, \eta^2_p = 0.032$). In the 200-ms displays, there was a minor effect of reporting method ($F(1, 2646) = 4.5, p = 0.05, \eta^2_p = 0.184$), and numerosity ($F(6, 2646) = 39.8, p < 0.001, \eta^2_p = 0.666$), with a minor interaction ($F(6, 2646) = 2.2, p = 0.05, \eta^2_p = 0.097$). See Figure 17. Since there was an effect of condition in the proportion of correct trials, pairwise comparisons were examined and indicate an advantage for the tally task in 200-ms displays with 8 items (MD = 0.11, $p < 0.05$) and 9 items (MD = 0.16, $p < 0.01$).

ANOVA results for the average miscounts (absolute value) in the 50-ms condition also indicate no effect of reporting method ($F(1, 2645) = 0.1, p = 0.95, \eta^2_p = 0.001$), but there was an effect of numerosity ($F(6, 2645) = 52.8, p < 0.001, \eta^2_p = 0.725$), without an interaction ($F(6, 2645) = 0.6, p = 0.75, \eta^2_p = 0.028$). In the 200-ms displays, the effect of reporting method was not significant ($F(1, 2645) = 3.0, p = 0.10, \eta^2_p = 0.129$), but there was a significant effect of numerosity ($F(6, 2645) = 33.7, p < 0.001, \eta^2_p = 0.628$), without an interaction ($F(6, 2645) = 1.1, p = 0.35, \eta^2_p = 0.053$). See Figure 18.

**Subitizing range**

The Scheffé test for multiple comparisons was conducted on the average miscounts in order to identify homogeneous subsets. The results from this test indicate the following homogeneous subsets in the tally task (control): for 50-ms displays, the numerosity range of 2-5 (average miscounts 0.01 to 0.08, $p = 0.83$); and for 200-ms displays, the range of 2-6 (average miscounts 0.01 to 0.10, $p = 0.16$). The following
subsets were identified in the pointing task: for 50-ms displays, the range of 2-5 (average miscounts 0.01 to 0.04, $p = 0.99$); and for 200-ms displays, the range of 2-6 (average miscounts 0.01 to 0.06, $p = 0.75$). These subsets were the same for proportion of correct trials. Additionally, a trend analysis tested for significant deviations from linearity for each duration and reporting method separately among the various numerosity ranges. In the control task, the largest range whose trend did not significantly deviate from linearity was the 2-6 range for both 50-ms displays ($\beta = 0.04, p < 0.01$) and 200-ms displays ($\beta = 0.02, p < 0.01$). In the pointing task, the last range without deviations from linearity was the 2-5 range for 50-ms displays ($\beta = 0.01, \text{N.S.}$) and the 2-6 range for 200-ms displays ($\beta = 0.03, p < 0.01$). The Bayesian test for flatness identified in the control task a subitizing range of 2-3 for 50-ms displays ($P = 0.975$) and 2-5 for 200-ms displays ($P = 0.90$). In the pointing task, this range was 2-5 for 50-ms displays ($P = 0.90$) and 2-4 for 200-ms displays ($P = 0.85$).

Finally, a simple linear function was computed to compare the numerosity response to the display numerosity. There was a higher degree of linear fit for the pointing task than the tally task in the “enhanced subitizing range” for 50-ms displays, but performance in 200-ms displays was similar (see summary of results in Table 5).

*Localization accuracy*

ANOVA results for the magnitude of location errors were similar to the results of previous experiments, indicating significant main effects for duration ($F(1,2378) = 6.4$, $p < 0.05$, $\eta^2_p = 0.224$) and numerosity ($F(6,2378) = 35.1, p < 0.001, \eta^2_p = 0.623$), with no interaction ($F(6,2378) = 1.2, p = 0.33, \eta^2_p = 0.051$). See Figure 19.
Differences between Experiments 3A & 3B

The numeral selection responses from Experiments 3A & 3B were tested for differences in performance. ANOVA results indicate that there were no significant differences between the two versions of the tallying task. ANOVA results for the proportion of correct trials showed no effect for experiment in the 50-ms condition ($F(1,2954) = 0.1, p = 0.76, \eta^2_p = 0.016$) and the 200-ms condition ($F(1,2954) = 0.5, p = 0.51, \eta^2_p = 0.075$). Similarly, ANOVA for the average miscounts showed no effect for experiment in the 50-ms condition ($F(1,2953) = 0.5, p = 0.53, \eta^2_p = 0.068$) and the 200-ms condition ($F(1,2952) = 0.4, p = 0.63, \eta^2_p = 0.001$).

<table>
<thead>
<tr>
<th>Task</th>
<th>Typical subitizing range</th>
<th>Enhanced subitizing range</th>
<th>Full range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-4 items</td>
<td>5-9 items</td>
<td>2-6 items</td>
</tr>
<tr>
<td><strong>Pointing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>$\beta=.981, r^2=.963$</td>
<td>$\beta=.852, r^2=.725$</td>
<td>$\beta=.980, r^2=.961$</td>
</tr>
<tr>
<td>200-ms</td>
<td>$\beta=.979, r^2=.958$</td>
<td>$\beta=.893, r^2=.798$</td>
<td>$\beta=.986, r^2=.971$</td>
</tr>
<tr>
<td><strong>Tally</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>$\beta=.865, r^2=.930$</td>
<td>$\beta=.859, r^2=.737$</td>
<td>$\beta=.973, r^2=.947$</td>
</tr>
<tr>
<td>200-ms</td>
<td>$\beta=.979, r^2=.958$</td>
<td>$\beta=.926, r^2=.858$</td>
<td>$\beta=.989, r^2=.978$</td>
</tr>
</tbody>
</table>

Table 5. Regression values when analyzing the correlation between the reported numerosity and the actual display numerosity in Experiment 3B. Note: All values reported are standardized $\beta$ and adjusted $r^2$; all significance levels were $p < 0.001$. 

Figure 17. Proportion of trials enumerated correctly for pointing and tally (row) report conditions in Experiment 3B (50-ms on left and 200-ms on right).

Figure 18. Average miscounts for pointing and tally (row) report conditions in Experiment 3B (50-ms on left and 200-ms on right)
Discussion

In contrast to Experiment 3A, there were no effects of reporting method within Experiment 3B. Enumeration performance in both the tallying procedure and the spatial marking procedure produced the same level of accuracy. This showed accurate recall in numerosities of up to 5 in the 50-ms displays and up to 6 in the 200-ms displays. The drop in performance in the pointing task was earlier (after five items) than in previous versions of this experiment. There was also a slight advantage for the tally task on 200-ms displays with 8-9 items. These differences may be due to the population of subjects or the use of a different computer to run the experiment.

Together with Experiment 3A, the results suggest an advantage for a nonverbal response that tallies counted items even when the tally does not correspond to an accurate location on the response screen. Performance in the control condition of Experiments 3A and 3B was better than the numeral selection tasks presented in Experiments 2A and 2B. To what extent is this result due to the nonverbal nature of the “clicking response” or to
the ability to “tally” already-counted items (e.g., in order to prevent miscounting by using an “external memory” to help keep track of counted items)? The next experiment addresses the question by examining whether or not persisting visual markers are required for this enhanced performance.

4.7. Experiment 4: Enumerating-by-Pointing vs. Invisible Tally

In Experiment 4, the nonverbal “clicking” nature of the response was isolated in order to examine enumeration performance when simply clicking a mouse to record perceived numerosity (with no visible marker to tally response). This experiment examined whether the pointing task benefits from a visual representation (i.e., tally) of already counted items or if the act of nonverbally reporting numerosity by simply clicking the mouse could enhance performance.

Method

Experiment 4 presented the same pointing task and stimuli as the previous experiments but implemented a variation in the control block. Here, observers were instructed simply to click the mouse pointer to record their numerosity response, so that each click corresponded to a single stimulus disc, but without the benefit of a consistently visible pointer or response markers to act as a sort of external memory (and aid localization or tallying). To ensure that a click registered, the mouse cursor would appear for 200 ms during each click. This control tested whether the act of “clicking” itself was responsible for increased numerosity performance, or if spatially-relevant markers were necessary to keep track of counted items (i.e., in the pointing task). Twenty-three Rutgers undergraduates completed both blocks in one 50-minute session for course credit.
Results

Enumeration accuracy

ANOVA were conducted on the proportion of trials with correct enumeration. In the control task (click only), enumeration accuracy as proportion of trials correctly enumerated showed main effects for both display duration \( (F(1,2898) = 24.2, p < 0.001, \eta^2_p = 0.524) \) and numerosity \( (F(6,2898) = 52.3, p < 0.001, \eta^2_p = 0.704) \), with an interaction \( (F(6,2898) = 7.4, p < 0.001, \eta^2_p = 0.252) \). For the pointing task, there were main effects for display duration \( (F(1,2898) = 9.0, p < 0.01, \eta^2_p = 0.291) \) and numerosity \( (F(6,2898) = 49.2, p < 0.001, \eta^2_p = 0.691) \), with an interaction \( (F(6,2898) = 3.6, p < 0.01, \eta^2_p = 0.141) \). See Figure E in the Appendix for charts displaying overall performance for individual subjects.

Again, the absolute values of the miscounts were analyzed using ANOVA methods (82% of miscounts in the control task were undercounts, and 87% of errors in the pointing task were undercounts). In the control task, there were main effects for display duration \( (F(1,2896) = 19.4, p < 0.01, \eta^2_p = 0.469) \) and numerosity \( (F(6,2896) = 28.3, p < 0.001, \eta^2_p = 0.563) \), with an interaction \( (F(6,2896) = 8.3, p < 0.001, \eta^2_p = 0.273) \). In the pointing task, there were effects for duration \( (F(1,2897) = 10.9, p < 0.01, \eta^2_p = 0.331) \) and numerosity \( (F(6,2897) = 44.9, p < 0.001, \eta^2_p = 0.671) \), with an interaction \( (F(6,2897) = 4.1, p < 0.001, \eta^2_p = 0.155) \).

Bonferroni results indicate that average miscounts in the control task were significantly worse in the 50-ms condition than the 200-ms condition for displays with 7 items \( (MD = 0.13, p < 0.01) \), 8 items \( (MD = 0.27, p < 0.01) \), and 9 items \( (MD = 0.25, p < 0.01) \). In the pointing task, the differences between the 50-ms and 200-ms conditions
were significant for displays with 6 items (MD = 0.07, \( p < 0.05 \)), 7 items (MD = 0.10, \( p < 0.05 \)), 8 items (MD = 0.16, \( p < 0.01 \)), and 9 items (MD = 0.36, \( p < 0.01 \)). These results were comparable in the proportion of correct trials measure.

**Enumeration accuracy: Differences in reporting method**

To examine how reporting method affected enumeration performance, ANOVAs were performed for the two display durations separately with reporting method and numerosity as factors. In the 50-ms displays, ANOVA on the proportion of correct trials showed a main effect for numerosity (\( F(6,2898) = 63.8, p < 0.001, \eta^2_p = 0.744 \)), but no effect of reporting method (\( F(1,2898) = 1.2, p = 0.29, \eta^2_p = 0.051 \)) and no interaction (\( F(6,2898) = 0.3, p = 0.94, \eta^2_p = 0.013 \)). In the 200-ms displays, there also was a main effect for numerosity (\( F(6,2898) = 48.1, p < 0.001, \eta^2_p = 0.686 \)), but no effect of reporting method (\( F(1,2898) = 0.1, p = 0.84, \eta^2_p = 0.002 \)) and no interaction (\( F(6,2898) = 1.1, p = 0.39, \eta^2_p = 0.046 \)). See Figure 20.

ANOVA for the average miscounts in the 50-ms condition also revealed a main effect for numerosity (\( F(6,2895) = 48.2, p < 0.001, \eta^2_p = 0.695 \)) but none for reporting method (\( F(1,2895) = 1.3, p = 0.27, \eta^2_p = 0.056 \)) and no interaction (\( F(6,2895) = 0.5, p = 0.83, \eta^2_p = 0.020 \)). In the 200-ms displays, there also was a main effect for numerosity (\( F(6,2898) = 37.6, p < 0.001, \eta^2_p = 0.631 \)), but no effect of reporting method (\( F(1,2898) = 0.2, p = 0.63, \eta^2_p = 0.011 \)) and no interaction (\( F(6,2898) = 0.6, p = 0.71, \eta^2_p = 0.028 \)). See Figure 21. Although there were no significant effects of reporting method, Bonferroni comparisons indicate a significant advantage for the pointing task (over the control task) in the 50-ms condition for displays with 2 items
(MD = 0.03, p < 0.01) and 9 items (MD = 0.12, p < 0.05). For the 200-ms displays, the advantage for the pointing task was found in 6-item displays (MD = 0.07, p < 0.05).

Subitizing range

The Scheffé test was used to identify homogeneous subsets in average miscounts. The results from this test indicate the following homogeneous subsets in the clicking task (control): for 50-ms displays, the numerosity range of 2-5 (average miscounts 0.01 to 0.07, p = 0.77); and for 200-ms displays, the range of 2-5 (average miscounts 0.01 to 0.07, p = 0.71). The following subsets were identified in the pointing task: for 50-ms displays, the range of 2-5 (average miscounts 0.01 to 0.06, p = 0.78); and for 200-ms displays, the range of 2-6 (average miscounts 0.01 to 0.07, p = 0.46). These subsets were also assessed using a trend analysis by testing for significant deviations from linearity (each duration and reporting method separately). In the control task, the largest range whose trend did not significantly deviate from linearity was the 2-4 range for 50-ms displays ($\beta = -0.01$, N.S.) and 2-6 range for 200-ms displays ($\beta = 0.03$, p < 0.01). In the pointing task, the last range without deviations from linearity were the 2-5 range for 50-ms displays ($\beta = 0.02$, p < 0.01) and the 2-6 range for 200-ms displays ($\beta = 0.02$, p < 0.01). The Bayesian test for flatness identified in the control task a subitizing range of 2-5 for 50-ms displays ($P = 0.85$) and 2-4 for 200-ms displays ($P = 0.90$). In the pointing task, this range was 2-4 for both 50-ms displays ($P = 0.95$) and 200-ms displays ($P = 0.99$).

A simple linear function again was computed to compare the numerosity response to the display numerosity. The degree of linear fit was higher for the pointing task than the control task, with a benefit for longer display durations. See Table 6 for these results.
Localization accuracy

ANOVA results for the magnitude of location errors were similar to the results of the previous Experiments, indicating main effects for display duration \( (F(1,2599) = 11.2, p < 0.01, \eta^2_p = 0.327) \) and numerosity \( (F(6,2599) = 40.9, p < 0.001, \eta^2_p = 0.635) \), with no interaction \( (F(6,2599) = 1.0, p = 0.42, \eta^2_p = 0.042) \). See Figure 22.

<table>
<thead>
<tr>
<th>Task</th>
<th>Typical subitizing range</th>
<th>Enhanced subitizing range</th>
<th>Full range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-4 items</td>
<td>5-9 items</td>
<td>2-6 items</td>
</tr>
<tr>
<td><strong>Pointing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>( \beta = .994, r^2 = .987 )</td>
<td>( \beta = .988, r^2 = .975 )</td>
<td>( \beta = .937, r^2 = .859 )</td>
</tr>
<tr>
<td>200-ms</td>
<td>( \beta = .995, r^2 = .989 )</td>
<td>( \beta = .919, r^2 = .844 )</td>
<td>( \beta = .993, r^2 = .985 )</td>
</tr>
<tr>
<td><strong>Click</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>( \beta = .987, r^2 = .975 )</td>
<td>( \beta = .852, r^2 = .726 )</td>
<td>( \beta = .973, r^2 = .946 )</td>
</tr>
<tr>
<td>200-ms</td>
<td>( \beta = .986, r^2 = .972 )</td>
<td>( \beta = .902, r^2 = .814 )</td>
<td>( \beta = .984, r^2 = .968 )</td>
</tr>
</tbody>
</table>

Table 6. Regression values when analyzing the correlation between the reported numerosity and the actual display numerosity in Experiment 4. Note: All values reported are standardized \( \beta \) and adjusted \( r^2 \); all significance levels were \( p < 0.001 \).

Figure 20. Proportion of trials enumerated correctly for pointing and invisible tally report conditions in Experiment 4 (50-ms on left and 200-ms on right).
Discussion

Experiment 4 indicates that there is a minor advantage when clicking a location with a visible marker than simply clicking (mainly for 200-ms displays with six items). This could be due to a simple recall error of keeping track of the number of clicks for larger numerosities in the control condition, since there were no constant visual markers.
to keep track of how many times the observer clicked the mouse to make the response. The subitizing range in this control, however, is lower than the “tally” control condition, which may be due to the “tagging” of already-counted items in the representation via an external memory that helps reduce miscounts. The next experiment examines whether a tagging benefit relies on persisting visual markers (as in Experiments 3A and 3B) or if the same benefit can be obtained simply by asking observers to visit the locations of the discs and click at that location without a persisting visual marker.

4.8. Experiment 5: Enumerating-by-Pointing vs. Invisible Location Markers

There are several possible explanations for the larger subitizing range observed with the pointing response in the previous experiments. 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 these experiments may tap into a different system of (motor) representation, and it is this location representation that allows accurate location marking, which in turn leads to external markings that could be used by the symbolic counting process. Experiment 5 examines whether pointing to the locations of the discs on the stimulus displays to record responses (without the aid of persisting markers) enhances performance than when simply clicking in Experiment 4. Here, observers marked locations of stimulus discs in two blocks, but the marker would not persist in the control block (i.e., disappear). This aims to address the question of whether simply pointing to locations provide an advantage during enumeration.
Method

Experiment 5 collected location information in both the test and control blocks. In the control block, which was always presented first, observers placed “X” markers on the screen locations for each object they saw, but the marker disappeared after 200 ms; otherwise, the design of this control block was identical to the standard pointing task in the second block (identical to previous experiments). This allowed testing whether or not locating objects without the aid of a stable visible location marker would affect enumeration and location accuracy. Twenty-one Rutgers undergraduates completed both blocks in one 50-minute session for course credit or payment.

Results

Enumeration accuracy

ANOVA were conducted on the proportion of trials with correct enumeration. In the control block with the temporary (200-ms) location markers, there were main effects for both duration ($F(1,2646) = 23.8, p < 0.001, \eta^2_p = 0.544$) and numerosity ($F(6,2646) = 69.6, p < 0.001, \eta^2_p = 0.777$), with an interaction ($F(6,2646) = 4.7, p < 0.001, \eta^2_p = 0.191$). In the pointing task, there were effects for duration ($F(1,2646) = 12.5, p < 0.01, \eta^2_p = 0.385$) and numerosity ($F(6,2646) = 40.8, p < 0.001, \eta^2_p = 0.671$), with an interaction ($F(6,2646) = 6.7, p < 0.001, \eta^2_p = 0.252$). See Figure F in the Appendix for charts displaying overall performance for individual subjects.

The absolute values of the miscounts were analyzed (88% of errors in the control task were undercounts, and 82% of errors in the pointing task were undercounts). In the control block, ANOVA results indicate main effects for both display duration ($F(1,2645) = 21.6, p < 0.001, \eta^2_p = 0.520$) and numerosity ($F(6,2645) = 42.1, p < 0.001$,}$
$eta^2_p = 0.678$), with an interaction ($F(6,2645) = 6.1, p < 0.001, eta^2_p = 0.234$). Also in the standard pointing task, there were effects for display duration ($F(1,2645) = 13.9, p < 0.01, eta^2_p = 0.410$) and numerosity ($F(6,2645) = 37.7, p < 0.001, eta^2_p = 0.653$), with an interaction ($F(6,2645) = 8.1, p < 0.001, eta^2_p = 0.288$).

Display duration had more of an effect on the numeral selection task than the pointing task. Bonferroni tests confirmed that the enumeration in the control task was significantly worse in the 50-ms condition than the 200-ms condition for displays with 2 items (MD = 0.03, $p < 0.05$), 7 items (MD = 0.16, $p < 0.01$), 8 items (MD = 0.24, $p < 0.01$), and 9 items (MD = 0.45, $p < 0.01$). In the pointing task, the differences between the 50-ms and 200-ms conditions were significant for displays with 2 items (MD = 0.04, $p < 0.05$), 8 items (MD = 0.23, $p < 0.01$), and 9 items (MD = 0.36, $p < 0.01$). These results were comparable in the proportion of correct trials measure.

*Enumeration accuracy: Differences in reporting method*

To examine how reporting method affected enumeration performance, ANOVAs were performed for the two display durations separately with reporting method and numerosity as factors. In the 50-ms displays, ANOVA on the proportion of correct trials showed a main effect for reporting method ($F(1,2646) = 6.2, p < 0.05, eta^2_p = 0.237$) and numerosity ($F(6,2646) = 78.3, p < 0.001, eta^2_p = 0.797$), with an interaction ($F(6,2646) = 2.4, p < 0.05, eta^2_p = 0.106$). In the 200-ms displays, there was no effect for reporting method ($F(1,2646) = 1.3, p = 0.26, eta^2_p = 0.062$), but there was a main effect for numerosity ($F(6,2646) = 38.5, p < 0.001, eta^2_p = 0.658$), with a minor interaction ($F(6,2646) = 2.0, p = 0.07, eta^2_p = 0.091$). See Figure 23.
ANOVAs for the average miscounts (absolute value) in the 50-ms condition revealed effects for reporting method ($F(1,2644) = 6.7, p < 0.05, \eta^2_p = 0.252$) and numerosity ($F(6,2644) = 56.7, p < 0.001, \eta^2_p = 0.739$), with an interaction ($F(6,2644) = 4.7, p < 0.001, \eta^2_p = 0.190$). In the 200-ms displays, there was no effect for reporting method ($F(1,2646) = 2.0, p = 0.17, \eta^2_p = 0.091$), but there was a main effect for numerosity ($F(6,2646) = 34.3, p < 0.001, \eta^2_p = 0.632$), with an interaction ($F(6,2646) = 3.4, p < 0.01, \eta^2_p = 0.147$). See Figure 24. Bonferroni comparisons of counting performance differences between reporting methods indicate a significant advantage for the standard pointing task (over the control task) in the 50-ms condition for displays with 6 items (MD = 0.07, $p < 0.01$), 7 items (MD = 0.20, $p < 0.01$), 8 items (MD = 0.20, $p < 0.01$), and 9 items (MD = 0.22, $p < 0.01$). In the 200-ms displays, there was an advantage for the control task in displays with 8 items (MD = 0.19, $p < 0.01$) and 9 items (MD = 0.13, $p < 0.05$). Again, the 50-ms displays were more susceptible to the different reporting methods than the 200-ms displays.

Subitizing range

The Scheffé test for multiple comparisons was performed to identify the homogeneous subsets. The results from this test indicate the following homogeneous subsets in the pointing with no marker task (control): for 50-ms displays, the numerosity range of 2-5 (average miscounts 0.02 to 0.10, $p = 0.69$); and for 200-ms displays, the range of 2-6 (average miscounts 0.01 to 0.11, $p = 0.08$). The following subsets were identified in the standard pointing task: for 50-ms displays, the range of 2-6 (average miscounts 0.01 to 0.10, $p = 0.18$); and for 200-ms displays, the range of 2-6 (average miscounts 0.01 to 0.10, $p = 0.07$). These subsets were the same for proportion of correct
trials, except 200-ms displays in the control condition only had a range of 2-5. A trend analysis tested for significant deviations from linearity for each duration and reporting method separately among the various numerosity ranges. In the control task, the largest range whose trend did not significantly deviate from linearity was the 2-5 range for 50-ms displays ($\beta = 0.02, p < 0.01$) and the 2-6 range in 200-ms displays ($\beta = 0.03, p < 0.01$). In the pointing task, the last range without deviations from linearity was the 2-6 range for both 50-ms ($\beta = 0.02, p < 0.01$) and 200-ms displays ($\beta = 0.02, p < 0.01$). The Bayesian test for flatness identified in the control task a subitizing range of 2-4 for 50-ms displays ($P = 0.99$) and 2-3 for 200-ms displays ($P = 0.99$). In the pointing task, this range was 2-5 for 50-ms displays ($P = 0.975$) and 2-4 for 200-ms displays ($P = 0.99$).

A simple linear function again was computed to compare the numerosity response to the display numerosity. The degree of linear fit for the “enhanced subitizing range” was almost identical for both the standard pointing response and the control response (with no marker). See Table 7 for a summary of these results.

**Localization accuracy**

ANOVA results for the magnitude of location errors in the standard pointing task were similar to the results of the previous Experiments, indicating main effects for display duration ($F(1,2365) = 5.2, p < 0.05, \eta^2_p = 0.175$) and numerosity ($F(6,2365) = 25.3, p < 0.001, \eta^2_p = 0.518$), but no interaction ($F(6,2365) = 1.1, p = 0.35, \eta^2_p = 0.049$). See Figure 25. In the control block where the observers marked the locations but the markers did not stay visible on the screen, there were no main effects for display duration ($F(1,2080) = 0.1, p = 0.94, \eta^2_p = 0.001$) or numerosity ($F(6,2080) = 1.5, p = 0.19, \eta^2_p = 0.061$), with no interaction ($F(6,2080) = 0.2, p = 0.98$,}
Not surprisingly, localization performance suffered without the benefit of stable spatial markers that could be used as landmarks for the pointing task.

The effect of reporting method on localization accuracy was examined by directly comparing the results from the two methods. ANOVA results for the 50-ms displays show a main effect for reporting method ($F(1,2198) = 4.8, p < 0.05, \eta^2_p = 0.181$) and numerosity ($F(6,2198) = 6.9, p < 0.001, \eta^2_p = 0.231$), with an interaction ($F(6,2198) = 3.4, p < 0.01, \eta^2_p = 0.135$). In the 200-ms displays, there was no effect for reporting method ($F(1,2247) = 0.3, p = 0.58, \eta^2_p = 0.015$), but there was an effect of numerosity ($F(6,2247) = 10.3, p < 0.001, \eta^2_p = 0.311$) with an interaction ($F(6,2247) = 11.2, p < 0.001, \eta^2_p = 0.324$). Bonferroni comparisons of localization differences between reporting methods indicate a significant advantage for the standard pointing task (over the control task) in the 50-ms condition for displays with 3 items (MD = 7.6, $p < 0.05$). For the 200-ms displays, there was an advantage for the pointing task in displays with 2 items (MD = 17.2, $p < 0.01$) and 3 items (MD = 14.5, $p < 0.01$).

<table>
<thead>
<tr>
<th>Task</th>
<th>Typical subitizing range</th>
<th>Enhanced subitizing range</th>
<th>Full range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-4 items</td>
<td>5-9 items</td>
<td>2-6 items</td>
</tr>
<tr>
<td>Pointing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>$\beta = 0.961, r^2 = 0.923$ : $\beta = 0.882, r^2 = 0.799$</td>
<td>$\beta = 0.983, r^2 = 0.967$ : $\beta = 0.601, r^2 = 0.360$</td>
<td>$\beta = 0.973, r^2 = 0.947$</td>
</tr>
<tr>
<td>200-ms</td>
<td>$\beta = 0.993, r^2 = 0.986$ : $\beta = 0.916, r^2 = 0.861$</td>
<td>$\beta = 0.991, r^2 = 0.982$ : $\beta = 0.739, r^2 = 0.545$</td>
<td>$\beta = 0.983, r^2 = 0.966$</td>
</tr>
<tr>
<td>Pointing (no marker)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-ms</td>
<td>$\beta = 0.971, r^2 = 0.944$ : $\beta = 0.820, r^2 = 0.690$</td>
<td>$\beta = 0.979, r^2 = 0.958$ : $\beta = 0.513, r^2 = 0.262$</td>
<td>$\beta = 0.959, r^2 = 0.919$</td>
</tr>
<tr>
<td>200-ms</td>
<td>$\beta = 0.986, r^2 = 0.972$ : $\beta = 0.878, r^2 = 0.794$</td>
<td>$\beta = 0.987, r^2 = 0.974$ : $\beta = 0.640, r^2 = 0.408$</td>
<td>$\beta = 0.974, r^2 = 0.948$</td>
</tr>
</tbody>
</table>

Table 7. Regression values when analyzing the correlation between the reported numerosity and the actual display numerosity in Experiment 5. Note: All values reported are standardized $\beta$ and adjusted $r^2$; all significance levels were $p < 0.001$. 

$\eta^2_p = 0.010$). See Figure 26.
Figure 23. Proportion of trials enumerated correctly for pointing task, in Experiment 5 with and without visible markers (50-ms on left and 200-ms on right).

Figure 24. Average miscalculation for pointing task, with and without visible markers in Experiment 5 (50-ms on left and 200-ms on right).
Figure 25. Average distance between stimulus-response pairs for Experiment 5 pointing task, in pixels (left y-axis) and visual degrees (right y-axis).

Figure 26. Average distance between stimulus-response pairs for Experiment 5 pointing task without visible markers.
Discussion

Experiment 5 indicates that there is an advantage in the standard enumerating-by-pointing task over the version where the markers disappear after a brief appearance. This affected enumeration performance and, less surprisingly, localization performance. The results in this control task, however, were still better than the numeral selection control in Experiments 2A and 2B. This suggests that visiting the locations of the stimulus discs allows the observer to better keep track of already-counted items as they make their responses. In the 50-ms displays, there was a small effect of reporting method on enumeration performance, suggesting that the persisting markers helped trials with very brief stimulus presentations (perhaps as a way to help reinforce the representation of the stimulus display). Localization errors were reduced for the block with persisting markers, an expected result; however, this advantage was only present in the displays with 2-4 discs and errors were comparable in displays with 5 or more discs. This suggests that the errors found in localizing displays with 6 or more items are influenced by a density saturation effect and could not exceed a certain number of pixels due to the small displays (i.e., number of discs interact with spatial dispersion).

4.9. Discussion for Experiments 1-5

The results from Experiments 1-5 suggest an interesting result: a non-cardinal and nonverbal report of numerosity either by tallying or pointing to the locations of individual discs (in a set of discs that needs to be counted) increases the subitizing limit to about six items. The various nonverbal reporting methods introduced in Experiments 3A, 3B, 4, and 5 produced better enumeration performance than the numeral selection task in
Experiments 2A and 2B. Table 8 summarizes the subitizing ranges identified by the trend analysis, the Scheffé test for flatness, and the Bayesian test for flatness in each of the Experiments described above. When responding by selecting a cardinal numeral, enumeration performance followed the “standard” result pattern, where enumeration accuracy declined after four items, instead of six which was observed in the pointing task (Experiments 2A and 2B). A smaller, but still significant, advantage can also be seen in another form of nonverbal response—tallying the number of items seen (Experiments 3A and 3B). Here, enumeration performance in the tally condition remained high (up to five items). When the tally process did not present a visible marker to help keep track of clicks (Experiment 4), performance was not as good as in the pointing task but better than when numerosity was reported by selecting a numeral symbol. Visiting locations with the cursor, however, and placing markers that disappeared after 200 ms produced the same advantage as the standard pointing task with visible markers (Experiment 5). Overall, the 200-ms displays produce more stable results with a subitizing range up to 5 or 6 items. The 50-ms displays, however, still produce the standard subitizing range of up to 4 items, with several occasions where this range was increased to 5 or 6 items.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Reporting Method</th>
<th>Duration</th>
<th>Trend analysis</th>
<th>Scheffé subsets</th>
<th>Bayesian analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>Pointing</td>
<td>50-ms</td>
<td>2-6</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>350-ms</td>
<td>2-7</td>
<td>2-6</td>
<td>2-4</td>
</tr>
<tr>
<td>Experiment 2A</td>
<td>Pointing</td>
<td>50-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td>Numeral selection (ring)</td>
<td>50-ms</td>
<td>2-4</td>
<td>2-4</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-5</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td>Experiment 2B</td>
<td>Pointing</td>
<td>50-ms</td>
<td>2-5</td>
<td>2-5</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-6</td>
</tr>
<tr>
<td></td>
<td>Numeral selection (row)</td>
<td>50-ms</td>
<td>2-5</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-4</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td>Experiment 3A</td>
<td>Pointing</td>
<td>50-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td>Tally (central grid)</td>
<td>50-ms</td>
<td>2-5</td>
<td>2-5</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-5</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td>Experiment 3B</td>
<td>Pointing</td>
<td>50-ms</td>
<td>2-5</td>
<td>2-5</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td>Tally (bottom row)</td>
<td>50-ms</td>
<td>2-6</td>
<td>2-5</td>
<td>2-3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-5</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>Pointing</td>
<td>50-ms</td>
<td>2-5</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td>Click only</td>
<td>50-ms</td>
<td>2-4</td>
<td>2-5</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>Pointing</td>
<td>50-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td>Pointing (no markers)</td>
<td>50-ms</td>
<td>2-5</td>
<td>2-5</td>
<td>2-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>200-ms</td>
<td>2-6</td>
<td>2-6</td>
<td>2-3</td>
</tr>
</tbody>
</table>

Table 8. Summary of the subitizing ranges observed in Experiments 1-5. The subitizing range was identified using the following methods: 1) linear trend analysis (as the largest range of numbers that did not deviate from a flat linear trend); 2) Scheffé test for multiple pairwise comparisons (as the homogeneous subsets with means that do not significantly differ); and 3) Bayesian null hypothesis testing (as the subsets with mean comparisons that favor accepting the null hypothesis).
4.10. Analysis of the Localization Data

In this section, data from the pointing tasks in Experiments 1-5 were combined to analyze the effect of spatial characteristics on enumeration and localization accuracy, and specifically to explore the nature of spatial memory during the pointing task. Do the relationships between the objects on the stimulus displays bias the representation of space (e.g., as described in Verbeek & Spetch, 2008)? Does the pointing task reveal a bias towards the central tendency—or prototype—in the observers’ representations of the configuration of the objects on the stimulus display (e.g., as described in Huttenlocher, et al., 1994; Spencer & Hund, 2002)? These sorts of questions will be addressed by examining the spatial relationships among the coordinates of the observers’ responses as well as the effects of regularity in spacing between the discs on stimulus displays. Since the current study used very brief presentations of the enumeration stimulus, observers must quickly extract the information from these masked displays in order to reproduce the locations of the discs on the display. The localization errors that emerge from this task can be examined to characterize the mechanism that is involved. Biases in the representations of the spatial configuration of the disc arrays can be identified by detecting the magnitude and direction of errors, as well as the compression of space.

Studies on the spatial encoding of object locations have shown that observers tend to remember locations by using spatial cues to categorize locations according to geometric “prototypes” (Huttenlocher, et al., 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, et al., 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). Similar studies also 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), which could impact enumeration.

The spatial encoding of object locations often exploits scene properties. Egocentric encoding of locations (i.e., in relation to the observer) is often used to help represent object locations (Shelton & McNamara, 2001). External landmarks, however, also have been shown to aid location memory (Lee, Shusterman, & Spelke, 2006; Oakes, Hurley, Ross-Sheehy, & Luck, in press). Therefore, instead of geocentric spatial encoding, the relationships between objects can bias the representation of space (Verbeek & Spetch, 2008). One question that arises from these various spatial representation strategies is the nature of the representation for objects that must be enumerated. How important is it to represent space accurately when enumerating (i.e., is the individuation of an object sufficient for enumeration, without an accurate memory of locations)?

Another useful approach for understanding enumeration abilities is to apply information theory principles in the study of visual perception. A study by Brady, Konkle, & Alvarez (2009), for example, 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. It was hypothesized that redundancies in the input information would allow more content to 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 (Brady, et al., 2009). Other studies also have shown that information can be compressed to optimize perception by reducing the encoding of redundant information and prioritizing the encoding of changes in perceptual information (Barlow, 2001). Redundancies in vision can take the form of contour or texture constancy, for example, and a change in contour or texture carries more information than constancy. It has been shown that the visual system uses regularities to encode information more efficiently at various stages in visual processing, including the level of the individual neuron as well as population of neurons, with minimal loss of information (Simoncelli & Olshausen, 2001).

To examine how information processing limitations could affect enumeration in the current study, localization results were examined in relation to display characteristics in order to describe the observers’ spatial representations of objects selected for enumeration. This was achieved by comparing the spatial relationships between the discs on a test display to those in the observers’ responses—essentially a comparison between a stimulus and its representation. Following information theoretic principles in perception (e.g., Attnave, 1954; Brady, et al., 2009), one testable prediction is that displays with more regularity allow more information to be encoded into working memory and lead to better enumeration and localization performance.

In the current study, redundancy may take the form of regularity in spacing between stimulus objects, where higher spatial regularity will require less encoding of spatial information. The only constraint on the otherwise random placement of stimulus discs was the minimum spacing between the objects (with the primary goal to eliminate
crowding). Since these discs were randomly placed, display regularity was obtained by applying Delaunay Triangulation methods to identify triangle “simplexes” using all the disc coordinates on a display as the vertices for these triangles (Kendall, 1989). Figure 27 illustrates the stimulus data (blue) and response data (red) as well as the results of the triangulation process (centroid locations are marked with an asterisk). This triangulation was applied to the elements on both the stimulus and response displays, and the average area, side lengths of the resulting triangles, and variability in these values 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.

To further characterize the spatial nature of the displays, the centroid of disc locations was determined for both stimulus and response data by computing the mean of all x-y coordinates on a display (stimulus and response displays were computed separately). Humans can estimate quickly 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). Since this is a form of statistical summary representation, the computation of this centroid estimate may prove to be crucial when representing individual locations. To examine how localization was related to the centroid, the distances between the centroid and each element on a display were computed. These values from the stimulus data were compared to the values from the response data in order to estimate variability and compression. These analyses originally were performed for Experiment 1 and reported in Haladjian, Singh, Pylyshyn, & Gallistel (2010).
The regularity measures described above may be used to develop a model that predicts enumeration and localization performance. The current analysis 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, including the possible role of the visual indexing mechanism.
Figure 27. Representative location data with the triangle simplexes drawn. Each chart represents data from a single trial; blue corresponds to the stimulus data and red corresponds to the response data.
Combined Results from Experiments 1-5

ANOVA were conducted on the performance measures using an aggregated data file when appropriate (i.e., cases were mean performance for each subject in each condition). In these within-subject analyses, subject ID was included as a random variable to account for between-subject variance and the last numerosity (9) was removed to control for end effects. The following analyses examine results from 152 participants.

[Note: When looking at localization errors (e.g., Figures 30-38), trials with unpaired discs were excluded from this analysis, which primarily occurred when displays were miscounted (27% of possible data points); all trials were used when examining the effects of regularity on enumeration performance.]

Differences between experiments

To determine if there was an effect of the experiment from which these data were obtained, an ANOVA on the aggregated file including Experiment as a factor was conducted (seven total Experiments: 1, 2A, 2B, 3A, 3B, 4, and 5). This analysis found no effect of Experiment for proportion of correctly enumerated trials ($F(6,2030) = 1.5$, $p = 0.22$, $\eta^2_p = 0.280$). For average miscounts, again there was no effect of Experiment ($F(6,2030) = 1.5$, $p = 0.23$, $\eta^2_p = 0.248$). This indicates that overall enumeration performance in the pointing task was similar among the seven experiments, even though different populations were used for each experiment and that a different control task preceded the pointing task in Experiments 2-5.

Enumeration Accuracy

The ANOVA results on the proportion of trials correctly enumerated (Figure 28) showed main effects for display duration ($F(2,1044) = 79.9$, $p < 0.001$, $\eta^2_p = 0.479$) and
numerosity \( (F(6,1044) = 233.2, p < 0.001, \eta^2_p = 0.479) \), with an interaction \( (F(12,1044) = 27.7, p < 0.001, \eta^2_p = 0.200) \). ANOVA on the average miscounts (Figure 29) also showed main effects for duration \( (F(2,1044) = 89.0, p < 0.001, \eta^2_p = 0.506) \) and numerosity \( (F(6,1044) = 186.3, p < 0.001, \eta^2_p = 0.432) \), with an interaction \( (F(12,1044) = 30.4, p < 0.001, \eta^2_p = 0.259) \). [Note: See Figure G in the Appendix for performance split by individual subject, and Figure H plots enumeration performance on a log-scale to show error patterns among the different numerosities.] Observers were more likely to under-count than over-count the stimulus displays. Among Experiments 1-5, there were counting errors in 20% of all trials in the pointing task. Of these trials, 86% were undercounts and 14% were over-counts. Over-counting was more likely to occur on displays with 2-5 items. Additionally, over-counting was more likely to occur in the control condition of Experiments 2-5, which suggests that the pointing task was effective at preventing double-counting. See Table 9 for a summary of counting errors.

A Bonferroni test for multiple pairwise comparisons (on the non-aggregated data set) indicates that the counting performance in the 50-ms condition was significantly worse than the 200-ms condition for displays with 2 items \( (MD = 0.01, p < 0.05) \), 3 items \( (MD = 0.01, p < 0.01) \), 6 items \( (MD = 0.06, p < 0.01) \), 7 items \( (MD = 0.11, p < 0.01) \), 8 items \( (MD = 0.24, p < 0.01) \), and 9 items \( (MD = 0.38, p < 0.01) \). The 350-ms condition is not reported since it was administered to a small subset of the participants (i.e., only 24 observers in Experiment 1).

Subitizing range

The Scheffé test for multiple comparisons also was performed on the average miscounts to identify homogeneous subsets in the numerosity conditions (each display
duration separately). The results from this test indicate the following homogeneous subsets in the overall results of the pointing task: for 50-ms displays, the numerosity range of 2-5 (average miscounts 0.02 to 0.05, $p = 0.53$); for 200-ms displays, the range of 2-5 (average miscounts 0.01 to 0.05, $p = 0.08$); and for 350-ms displays, the range of 2-6 (average miscounts 0.01 to 0.08, $p = 0.27$).

These subsets were also identified using a trend analysis (for each duration separately) that tested for significant deviations from linearity in the following numerosity ranges: 2-4, 2-5, 2-6, 2-7, 2-8, and 2-9. When examining the average number of miscounts in the current study, the last group that did not deviate from linearity for 50-ms displays was the 2-5 numerosity range ($\beta = 0.01, p < 0.01$). The last group for the 350-ms displays that did not deviate from linearity was the 2-6 range ($\beta = 0.03, p < 0.01$). In the 200-ms displays, however, there were no groups that not deviate from linearity (for the 2-3 range, the slope is $\beta = 0.01, p < 0.01$). This unexpected result requires further examination, since the trend analyses for each experiment individually indicate a subitizing range of 2-6 items in the 200-ms condition of the pointing task. A similar pattern was also observed in the results from the Bayesian test for flatness, which identified a subitizing range of 2-4 for 50-ms displays ($P = 0.90$), only 2-3 for 200-ms displays ($P = 0.99$), and 2-4 for the 350-ms displays ($P = 0.90$).

Additionally, a simple linear function was computed to compare the numerosity response to the display numerosity. As Table 10 indicates, the degree of linear fit between display numerosity and numerical responses was comparably high for both the “standard subitizing range” (2-4 item displays) and the “extended subitizing range” (2-6 item displays).
Summary of Miscounts in the Pointing Task, by Numerosity Conditions for Each Trial

<table>
<thead>
<tr>
<th>#</th>
<th>&lt; -2</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>&gt; 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.0% (0)</td>
<td>0.0% (0)</td>
<td>0.6% (22)</td>
<td>98.8% (3396)</td>
<td>0.4% (15)</td>
<td>0.1% (3)</td>
<td>0.0% (0)</td>
</tr>
<tr>
<td>3</td>
<td>0.0% (0)</td>
<td>0.0% (1)</td>
<td>0.1% (5)</td>
<td>99.3% (3411)</td>
<td>0.5% (17)</td>
<td>0.0% (1)</td>
<td>0.0% (1)</td>
</tr>
<tr>
<td>4</td>
<td>0.0% (1)</td>
<td>0.0% (0)</td>
<td>0.5% (16)</td>
<td>97.6% (3354)</td>
<td>1.8% (62)</td>
<td>0.1% (2)</td>
<td>0.0% (1)</td>
</tr>
<tr>
<td>5</td>
<td>0.1% (2)</td>
<td>0.0% (1)</td>
<td>1.0% (34)</td>
<td>95.4% (3278)</td>
<td>3.3% (115)</td>
<td>0.2% (6)</td>
<td>0.0% (0)</td>
</tr>
<tr>
<td>6</td>
<td>0.1% (4)</td>
<td>0.1% (5)</td>
<td>5.2% (177)</td>
<td>89.8% (3085)</td>
<td>4.7% (160)</td>
<td>0.1% (4)</td>
<td>0.0% (1)</td>
</tr>
<tr>
<td>7</td>
<td>0.2% (8)</td>
<td>0.8% (29)</td>
<td>19.2% (661)</td>
<td>74.9% (2573)</td>
<td>4.6% (158)</td>
<td>0.2% (7)</td>
<td>0.0% (0)</td>
</tr>
<tr>
<td>8</td>
<td>0.2% (8)</td>
<td>5.3% (183)</td>
<td>37.2% (1279)</td>
<td>53.5% (1839)</td>
<td>3.6% (123)</td>
<td>0.1% (2)</td>
<td>0.1% (2)</td>
</tr>
<tr>
<td>9</td>
<td>2.4% (84)</td>
<td>17.4% (597)</td>
<td>47.5% (1632)</td>
<td>30.3% (1040)</td>
<td>2.2% (74)</td>
<td>0.2% (7)</td>
<td>0.1% (2)</td>
</tr>
<tr>
<td>Total</td>
<td>0.4% (107)</td>
<td>3.0% (816)</td>
<td>13.9% (3826)</td>
<td>79.9% (21976)</td>
<td>2.6% (724)</td>
<td>0.1% (32)</td>
<td>0.0% (7)</td>
</tr>
</tbody>
</table>

Table 9. Counting errors in the pointing task for Experiments 1-5 (n = 152). Data presented as the percent of trials having the magnitude of miscount (columns) with the number of trials in parentheses; each numerosity condition is on a separate row. Miscounts over or under 2 were combined in this summary chart since these errors occurred infrequently (< 0.5% of trials).

<table>
<thead>
<tr>
<th>Typical subitizing range</th>
<th>Enhanced subitizing range</th>
<th>Full range</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-4 items</td>
<td>5-9 items</td>
<td>2-6 items</td>
</tr>
<tr>
<td>50-ms</td>
<td>β= .981, r²= .961</td>
<td>β= .984, r²= .968</td>
</tr>
<tr>
<td>200-ms</td>
<td>β= .991, r²= .982</td>
<td>β= .991, r²= .983</td>
</tr>
<tr>
<td>350-ms</td>
<td>β= .990, r²= .980</td>
<td>β= .992, r²= .983</td>
</tr>
<tr>
<td>Combined</td>
<td>β= .986, r²= .972</td>
<td>β= .823, r²= .677</td>
</tr>
</tbody>
</table>

Table 10. Regression values when analyzing the correlation between the reported numerosity and the actual display numerosity for Experiments 1-5. Note: All values reported are standardized β and adjusted r²; all significance levels were p < 0.001.
Figure 28. Proportion of trials enumerated correctly in the pointing task in Experiments 1-5 (n = 152).

Figure 29. Average miscounts (absolute value) in the pointing task in Experiments 1-5 (n = 152).
Localization accuracy

An ANOVA with Experiment as a factor (for 50-ms and 200-ms displays, aggregated data set) indicates a significant effect of Experiment on localization errors \((F(6,2026) = 11.1, p < 0.01, \eta^2_p = 0.928)\). Bonferroni pairwise comparisons of mean performance indicate that Experiment 1 was significantly better than all the other experiments in the 200-ms display condition and from Experiment 3A in the 50-ms condition. Additionally, Experiment 3A showed lower performance than Experiments 2B and 3B. When running this ANOVA without Experiment 1, the effect of Experiment is no longer significant \((F(5,1704) = 7.2, p = 0.06, \eta^2_p = 0.919)\). This performance difference in Experiment 1 may be due to the greater number of trials per condition (although this was only 12 instead of 10 trials).

Overall ANOVA results (all Experiments) indicate main effects for display duration \((F(2,1040) = 39.4, p < 0.001, \eta^2_p = 0.312)\) and numerosity \((F(6,1040) = 124.4, p < 0.001, \eta^2_p = 0.281)\), with an interaction \((F(12,1040) = 3.0, p < 0.001, \eta^2_p = 0.034)\). Errors increased with larger numerosities and generally decreased with longer display durations (see Figure 30). Using the breakpoints identified in a trend analysis on the localization errors, the slope (unstandardized \(\beta\)) of these errors for the range of 2-7 items in 50-ms displays was \(\beta = 5.5\) pixels (constant = 47.5, \(p < 0.01\)), for the range of 2-7 items in 200-ms displays was \(\beta = 6.6\) (constant = 37.5, \(p < 0.01\)), and for the range of 2-8 items in the 350-ms displays was \(\beta = 5.3\) (constant = 33.5, \(p < 0.01\)).

Localization errors are plotted in Figure 31 as the Cartesian joint distribution of errors for each click number separately (50-ms and 200-ms durations combined). In this chart, the point of origin is a stimulus disc and error is the distance between this stimulus...
disc and the paired response marker. This figure illustrates the increasing trend in errors with each click. It must be noted, however, that the apparent decrease in the distribution of errors for the last two numerosities may be a function of there being fewer trials where 7 or 8 clicks were made on a screen, or a function of the density of the discs on a stimulus display (i.e., the maximum magnitude of error is limited by the increased density in the distribution of the stimulus discs). ANOVA with response number as a factor indicates a significant but small effect of response number ($F(7,97051) = 311.3, p < 0.001$, $eta_p^2 = 0.022$), which interacts with numerosity ($F(23,97051) = 21.0, p < 0.001$, $eta_p^2 = 0.005$) and display duration ($F(14,97051) = 3.7, p < 0.001$, $eta_p^2 = 0.001$).

It should be noted that this trend in localization error may be due to a cumulative effect, where error increases with each additional response. To depict the errors made in the first localization response, which was not affected by errors from previous clicks, the Cartesian joint distribution of errors for the first click is plotted in Figure 32 (for each numerosity, 50-ms and 200-ms combined). This chart illustrates a concentration of first-click errors near the point of origin for smaller numerosity displays that tends to expand for larger numerosities. Similar to previous analyses, Figure 33 plots localization errors as a function of numerosity but only for the first response. ANOVA results on first-click localization errors indicate significant effects of duration ($F(2,18982) = 45.8, p < 0.001$, $eta_p^2 = 0.360$) and numerosity ($F(6,18982) = 29.9, p < 0.001$, $eta_p^2 = 0.099$), with an interaction ($F(12,18982) = 2.0, p < 0.05$, $eta_p^2 = 0.025$). This confirms an effect of duration and numerosity for the first click, indicating that higher errors for displays with larger numerosities and shorter presentation durations are not only a cumulative effect of error that increases with each click (although this contributes to the overall results). There
is an underlying difference in spatial accuracy that is better for displays with low numerosities that are viewed longer, which affects even the first click response.

Finally, errors in relation to the point of fixation were examined. Each stimulus disc was assigned to one of four groups based on proximity to the fixation cross (in 100 pixel increments). Figure 34 plots the joint distribution of errors for these responses and indicates that localization errors increase when items are presented further from the point of fixation and into the periphery. ANOVA results (on the first response) confirm this result of increasing errors when discs are further from fixation. There was a significant effect of proximity for 50-ms \((F(3,10028) = 697.8, p < 0.001, \eta^2_p = 0.173)\) and 200-ms displays \((F(3,10685) = 553.1, p < 0.001, \eta^2_p = 0.135)\), but there were no interactions with display numerosity. Not surprisingly, the magnitude of localization errors increase as the distance from the central fixation cross increase (e.g., fovea effect), but this does not vary among the numerosity conditions.

Figure 30. Average distance between stimulus-response pairs, in pixels (left y-axis) and visual angle in degrees (right y-axis). For pointing task in Experiments 1-5 (n = 152). Note: 350-ms condition only in Experiment 1 (n = 24)
Figure 31. Cartesian joint errors, by order of response (click number). The “n” value on the plots represents the total number of response pairs plotted (all trials and all subjects). Experiments 1-5, 50-ms & 200-ms durations combined (n = 152).
Figure 32. Cartesian joint errors for the first-click response, split by numerosity condition. These plots illustrate the spread of localization errors, which tends to increase as the number of stimulus objects increase. The “n” value on the plots represents the total number of objects plotted (for all trials and all subjects).

Experiments 1-5, 50-ms & 200-ms durations combined (n = 152).
Figure 33. Average distance between stimulus-response pairs for the first click (Experiments 1-5; n = 152. Note: 350-ms condition only in Experiment 1; n = 24.)

Figure 34. Cartesian joint errors for first-click response, by distance from fixation (50-ms & 200-ms displays). This figure illustrates the increasing error distance as discs appeared further from the center of the screen. Experiments 1-5 (n = 152).

Distance from fixation is equal to polar distance values of:
1 = 0-99 pixels; 2 = 100-199 pixels; 3 = 200-299 pixels; 4 = 300+.
Direction and Magnitude of Localization Errors

The magnitude of localization errors is apparent from the results of the previous section. Primarily, the first click error increases for displays with shorter presentation durations and larger numerosities. In this section, the direction of these errors will be examined. To characterize this potential bias, the sine and cosine of the Euclidean relationship between the stimulus disc and response marker were computed, where the cosine value corresponds to the error on the x-axis (X-stimulus – X-response / total error distance) and the sine value corresponds to the error on the y-axis (Y-stimulus – Y-response / total error distance). The results for all responses in each trial are plotted in Figure 36 (split by duration). The blue crosses on the circle plot represent the direction of individual responses. The red crosses represent the resultant vector for each observer and characterizes the direction and magnitude of response bias (computed by adding all the
values for each observer). This illustrates an overall negative bias for responses on both the x- and y-coordinates, with a stronger bias in the 50-ms displays.

Figure 37 shows this directional bias for each of the stimulus display numerosities separately. Larger numerosities show a stronger negative bias, which indicates responses are made below and to the right of a stimulus disc. This trend is present in the first-click response as well, shown in Figure 38. Also evident in these figures is the increasing variability in the response bias for larger numerosities. Again, these charts show data from trials where stimulus discs and response markers were matched with high confidence, so these results illustrate the response bias present in the observers’ representations of the enumeration displays.

Figure 36. Circle plots of X (e-cos) Y (e-sin) errors for the first-click response, separated by stimuli display duration. The red crosses indicate the average “resultant vector” for each subject. The “n” value on the plots represents the total number of objects plotted. For pointing task in Experiments 1-5 (n = 152 for 50-ms and 200-ms displays; n = 24 for 350-ms displays).
Figure 37. Circle plot representing direction of errors (all responses), by number of stimuli objects shown; red crosses represent each subject’s average bias. For pointing task in Experiments 1-5, 50-ms & 200-ms durations combined (n = 152).
Figure 38. Circle plot representing direction of errors for the first-click response on each trial; red crosses represent each subject’s average bias for this first click response. For pointing task in Experiments 1-5, 50-ms & 200-ms durations combined (n = 152).
**Compression Around the Centroid**

The centroid (or center-of-mass) for each display was computed by calculating the mean $x$ and $y$-coordinates of all the discs on a display. Figure 30 plots the average distance of a stimulus disc from its centroid and a response marker from its centroid. This chart shows an underestimation in the distances between discs by approximately $1^\circ$ visual angle, and this sort of “compression” is centered around the centroid. The magnitude of compression was computed by taking the average distance from the centroid on a stimulus display and subtracting the average distance from the centroid on a response display. There were significant, but small, differences in the average magnitude of compression between the 50-ms (mean = 35.1 pixels) and 200-ms (mean = 33.0 pixels) conditions. [ANOVA results indicate significant effects of duration ($F(1,87971) = 37.3$, $p < 0.001$, $\eta^2 = 0.001$) and numerosity ($F(6,87971) = 145.9$, $p < 0.001$, $\eta^2 = 0.010$), with an interaction ($F(6,87971) = 6.6$, $p < 0.01$, $\eta^2 = 0.001$).]

Another way to characterize this compression is illustrated in Figure 40, which depicts the average segment lengths of the Delaunay Triangulation simplexes (shared sides were only counted once). These results also suggest a compression of the spatial representations by $\sim 1^\circ$, which decreases with larger numerosities (likely due to density). This compression can also be seen when examining the size of the maximal enclosing circle (i.e., the circle that surrounds all the items on a display). This captures the extreme values since the maximum enclosing circle is sensitive to the dispersion of the discs it must enclose. The average disc dispersion when measured as the radius of the minimum enclosing circle on a stimulus display ranged from 198 pixels (SD = 73) in 2-numerosity displays to 358 pixels (SD = 18) in 9-numerosity displays. For the response data, this
dispersion ranged from 174 pixels (SD = 73) to 301 pixels (SD = 44). This suggests a compression of 25 to 50 pixels (~0.75° to 1.5°) is present in the observers’ responses. These results support previous studies that have found representations of space to be compressed, that is, observers tend to underestimate the distances between items when recalling this information (e.g., Sheth & Shimojo, 2001). Furthermore, a follow-up experiment that examined eye-fixations during the pointing task showed that making more fixations during the response stage produced higher localization errors in the form of compression around the centroid (reported in Harman, Haladjian, Aks, & Pylyshyn, 2011). This supports studies that indicate spatial compression may be a result of saccades (e.g., Burr, Ross, Binda, & Morrone, in press).

Figure 39. Average centroid-to-disc distance in pixels in the pointing task for Experiments 1-5 (50-ms displays on right, 200-ms displays on left).
Effects of Display Regularity

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 maximum enclosing “circumcircle” around the display elements (to estimate disc dispersion). Since the 350-ms display condition was only administered in Experiment 1, only data from the 50-ms and 200-ms display durations are reported in the following section.

For calculating pattern regularity on a display, the mean and variance values were computed for the areas of triangle simplexes (identified by the Delaunay Triangulation), lengths of the connecting edges, and the radii of the maximal circles that circumscribe the triangle simplexes. 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. Only the effects of regularity as measured by the variability in the edge lengths
of the Delaunay triangle simplexes are reported here; however, similar patterns of results were obtained with the area of the simplexes and the size of the maximal circles (omitted from this report to reduce redundancy).

To compare levels of display regularity, the standard deviation of the triangle segments in 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 (Figure 41) and counting errors (Figure 42) as functions of increasing variability (decreasing regularity). 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 extremely well and there were no effects of regularity). There was a significant effect of regularity in localization performance for all the different numerosities displayed in Figure 41 (i.e., displays with 6-9 items in both 50-ms and 200-ms conditions). Additionally, there was a significant effect of regularity in enumeration performance for displays with 7-9 items (see Figure 42).

To compare the regularity of the stimulus and response patterns, the mean and standard deviation of the Delaunay triangle simplexes were examined. The original stimulus displays with 4-9 items had an average mean length of 335.4 pixels (SD = 34.9), while the simplex lengths in the responses averaged 291.3 pixels (SD = 33.1); these differences were significantly different (p < 0.01). This suggests that observers imposed regularity on the response patterns that was not present in the stimulus patterns and also supports the possibility that more “prototypical” shapes are applied to patterns in memory. This comparison also was performed on the response data that was
corrected by the Procrustes analysis. As noted earlier, a Procrustes analysis produces values that indicate how transformations occurred in a given shape when compared to the reference shape (i.e., in this case, the stimulus displays). By applying these values to produce “better fitting” response coordinates that maintained key elements of the shapes, such as the angles of the vertices in the Delaunay simplex triangles. The variance in the simplex segment lengths for these “uncompressed” response patterns was then found to be the same as the variance in the corresponding stimulus patterns; the Procrustes-corrected response data had an overall mean of 331.4 pixels, SD = 35.7. This indicates that when controlling for the observed compression, the distances between items did not differ significantly from the original displays.

Figure 41. Mean localization error as a function of original display variability as defined by the lengths of triangle simplexes (1 = less variable displays, 4 = most variable displays); 50-ms displays on right, 200-ms displays on left.
Figure 42. Mean counting errors (absolute miscounts) as a function of original display variability as defined by the lengths of triangle simplexes (1 = less variable displays, 4 = most variable displays); 50-ms displays on right, 200-ms displays on left.

Coefficient of variation

A signature of the nonverbal magnitude estimation mechanism is a scalar variability that produces a constant coefficient of variation (the ratio between the standard deviation and mean of observer responses). Figures 43-45 plot the results from the enumeration task, with each figure displaying results from one of the display duration conditions. The top chart plots the mean numerosity judgment as a function of display numerosity (blue line), with a lighter blue diagonal indicating the line of perfect enumeration performance. The standard deviation observed in these responses is plotted on the right y-axis (black line). The bottom chart is the resulting coefficient of variation (standard deviation divided by the mean response). These charts do not show a pattern of scalar variability; variability appears to increase in a manner that is not proportional to the mean response. Therefore, the coefficient of variance plotted in the bottom chart is not flat as would be expected when responses follow scalar variability. Note the performance difference for displays with three items in the 200-ms and 350-ms displays:
the low variability may indicate a grouping of “triplets” that enhances performance (e.g., as seen in Mandler & Shebo, 1982)

Figure 43. Coefficient of variance for the pointing task on 50-ms displays (n = 152).

Figure 44. Coefficient of variance for the pointing task on 200-ms displays (n = 152).
Figure 45. Coefficient of variance for the pointing task on 350-ms displays (only administered in Experiment 1; n = 24).

Discussion

The goal of this focused analysis on the spatial component of the pointing task was to characterize the mechanism that supports small-set enumeration by applying various theories of perceptual processing. For example, 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). This information theoretic account of perception is supported by the current results showing better performance on stimulus displays with more regular patterns and suggests that spacing regularity allows a more efficient encoding of object locations. When the triangle simplexes of a display have less variance, observers are more accurate in representing these displays and exhibit better enumerating and localization performance. Additionally,
there appears to be a tendency for compressing distances around the centroid, and 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 and towards more “prototypical” representations of shape (Huttenlocher, et al., 1994; Spencer & Hund, 2002). This observed tendency to impose regularity on variable displays supports findings from previous studies (e.g., Taylor, 1961), including those examining bias in representations of shape.

Increasing the stimulus exposure durations produced more accurate enumeration for displays with numerosities of 6-9 and more accurate location encoding for displays with 2-6 items even when localization error is examined for the first click to factor out the cumulative effect of response errors. 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 the individuation stage for the purpose of enumeration occurs independent of location-encoding: serial attention may be required to effectively encode locations while subitizing may be preattentive. This supports a visual indexing account 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 representations that include richer feature information (Pylyshyn, 1989).

The consistent finding (in Figures 30 and 31) that the longer the displays are exposed, the more accurate are the location responses is also compatible with a general finding in computational vision, where an efficient way of recognizing shapes is to use a
coarse representation and then to move to finer and finer scales (e.g., Amit, Geman, & Fan, 2004; Gangaputra & Geman, 2006). The underlying basis for this has been given in terms of a dynamic programming model by Raphael (2001). In fact, the lesson from some vision models is that localization of certain features may increase with time. These findings have been applied to cortical models of location and shape encoding. The results from the current study support this conceptualization of visual processing.

Although the current results suggest that the indexing mechanism is implemented for smaller numerosities, further experiments to support this conclusion are required since performance was at ceiling for smaller numerosities. Additionally, allowing observers to enumerate by location may be a more accurate demonstration of selection abilities during fast enumeration, and this 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 present in visual scenes. Further analyses of these location data from an information theoretic perspective promise to further our understanding about the spatial nature of numerosity representations.
Chapter 5. General Discussion

This study introduced an indirect way of determining how many briefly-presented items can be individuated and selected (or indexed) for further processing. In this “enumerating-by-pointing” task, observers indicated the location of each item in a set of 2-9 discs that were displayed briefly and masked. This design controlled exposure to the stimulus and prevented counting of persisting stimulus displays or after-images. By asking observers to indicate *where* each disc had been located, they could recall six items in over 90% of the trials. This capacity is in contrast to that obtained when observers 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 (as reviewed in Trick & Pylyshyn, 1994b). In the pointing task, subitizing was measured by the number of discrete markers placed on the response screen, which also produced a measure of localization accuracy. Localization was better for smaller numerosity displays and for displays that were presented for longer durations. These errors were systematic and were compressed around the center of mass of the stimulus discs. Additionally, both enumeration and localization were sensitive to spatial regularity, which suggests that errors may be introduced during the encoding and decoding of these representations in working memory.

*Subitizing limit*

Using traditional methods to examine changes in enumeration accuracy (i.e., linear trend analysis), results from Experiment 1 indicate a subitizing limit of up to six items with 50-ms and 200-ms displays, and up to seven items with 350-ms displays. This range was confirmed in Experiments 2-5, which compared the pointing task to various
controls. The controls produced inferior or comparable performance, depending on the control task, with a subitizing range of four or five items in the 50-ms conditions and five or six items in the 200-ms conditions (Table 8 summarizes the subitizing ranges). Overall, these ranges are comparable or slightly higher than those reported in previous subitizing studies. When examining this range using more a stringent Bayesian null hypothesis test for flatness (Gallistel, 2009), the subitizing ranges are slightly smaller, with a subitizing range of 2-5 for all durations of the pointing task, which is better than that of the numeral selection task (2-4 range). This suggests that a closer examination of the advantages for the different methods that can be used to determine the subitizing range is warranted. For the current study, discrepancies in the subitizing ranges should be kept in mind when interpreting the results. Nevertheless, there is a consistent trend that indicates a higher subitizing range for the pointing method (over the numeral report method) regardless of the computation that produces this range.

If the subitizing limit is taken to be the largest number of individuals that are indexed under ideal perceptual conditions (i.e., longer presentation durations), then the results from these experiments indicate that this number is around six items. This limit, however, may also be taken to be the largest number of items that can be retained under less favorable viewing conditions, for example, in the 50-ms display condition. The results indicate 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. Presentation durations of only 50-ms may limit the earliest individuation stage, which is only one part of the subitizing process, and such brief presentations result in less
stable representations (as suggested by Mandler & Shebo, 1982). This result may also indicate an indexing capacity of four or five items, which can be increased on 200-ms displays through grouping strategies or the deployment of (a limited) serial attention.

The observed ability to enumerate accurately at least five objects even when the stimuli were viewed briefly (50-ms) challenge the results from a recent study conducted by Poiese et al. (2008). When observers were presented a masked stimuli for 50-83 ms, enumeration accuracy was reduced in the subitizing range. They took this result as an indication that a serial processing of small numerosities was required to enumerate small-set displays because a parallel process should not have been affected by this brief display. The results from the current study, however, show high enumeration accuracy for at least four or five items on 50-ms displays in any of the reporting methods. The discrepancy in performance between these two studies is likely due to the nature of the subitizing task in the Poiese et al. study, which included a visual search task where observers were required to report the number of lines that were slanted in a certain direction among distractors that were slanted in the opposite direction. To achieve this, the visual system must search through the array of items in order to determine which items to count. This additional step of discrimination requires serial attention, and it is this stage that suffers during limited processing durations. In contrast, the pointing task tested in the current study simply asked observers to report the number of items on the screen without the need to discriminate targets from nontargets; this provides a more accurate measure of individuation and subitizing.

The results from the current study also contrast with the results from other nonverbal enumeration studies. As previously described, Cordes et al. (2001) compared
response performance on verbal and nonverbal counting tasks and found that
enumeration performance exhibited scalar variability in the nonverbal counting condition. This result was taken as evidence that a single magnitude mechanism generates the representations of small and large numerosities in nonverbal counting tasks. In the current study, the pointing task can be considered a nonverbal report of numerosity since verbal counting was prevented on the brief stimulus displays and the clicking responses did not require the assignment of any cardinal labels. The results from this nonverbal pointing task, however, did not produce scalar variability (Figures 43-45). This may be taken as evidence against the implementation of the magnitude estimation mechanism for producing the numerosity response and instead relies on the individuation mechanism, which may be emphasized by the “pointing” nature of the task.

The use of motor “pointing” gestures may explain the better performance in the pointing task over the cardinal response. 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 the current study may tap into a different system of (motor) representation, which in turn leads to external markings (e.g., markings in a proprioceptive space) that could be used by the symbolic counting process. This explanation is supported by the results from Experiments 2A and 2B that compare performance between the pointing task and a cardinal numeral selection task. A somewhat surprising result is found in Experiments 3A and 3B, where enumeration is enhanced, though to a lesser degree, even when the response was a spatially-irrelevant “tally”. Here, performance is better than the numeral selection method but not as good as the pointing method. This may indicate that the
nonverbal nature of the response methods contributes to the observed increase in accuracy, although it may still depend on the visual feedback from the tally marks which provide a memory aid as to how many items had been taken into account. The observer may be able to put the tally marks into one-one correspondence with individual disks, even though the unique identifying information provided by correctly-located tally marks is absent.

Another possible account for the difference between the pointing and cardinal reporting methods relies on Visual Indexing Theory. This theory (Pylyshyn, 2003, 2007) proposes a limited number of indexes each of which is automatically captured or activated by 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. To derive the symbolic cardinality of the set of indexed objects, a subsequent stage of enumeration is required. According to this account, when there are fewer objects than the indexing limit, enumeration may operate over active indexes 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 enumeration performance of small and large sets is affected differently by temporal and spatial factors, since the duration of the stimulus display may affect the first (indexing) stage while other factors, such as set size, may affect the second (counting) stage. Thus it is consistent with the observation from Experiment 1 that the “knee” or inflection of the performance curves (Figures 2 and 3) appears to shift and indicates higher performance as the stimulus
duration increases. This also applies to the results in Experiments 2A and 2B, 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).

Another account for why the pointing task produces a higher subitizing capacity 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 explain 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, providing that keeping track of which items were counted benefits from the motor representation (e.g., via the dorsal visual stream). If the objects are no longer present, as in the current study, using this motor representation to place marks on their former locations can help keep track of already-counted objects. As long as the observer 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 provide a visible tally that can be used to identify already-counted objects. Also, there may be “inhibitory” processes occurring when observers are marking locations on the screen (e.g., Tipper, Driver, & Weaver, 1991; Watson & Humphreys, 1997), which helps prevent repeated counting of the same item. That is, by “acting” on the display, the execution of motor actions in correspondence with spatial attention could result in inhibitory effects on already-visited spatial regions.

This “marking” explanation can account for the performance differences in Experiments 2A and 2B. However, it is not obvious how this account applies to
Experiments 3A and 3B, where observers used a spatially-irrelevant tally to report numerosity and also showed a slight increase in subitizing performance over the numeral selection task. It may be that one “tally” on the screen can correspond to one disc in the observer’s representation of the stimulus. By providing a visible tally on the response screen, the observer is able to make a nonverbal correspondence with the items that are being counted. Therefore, this sort of “external memory” may be responsible for the increase in the subitizing range. Indeed, some have argued that external cues can serve to minimize memory resources in perception (e.g., O'Regan, 1992), and an external tally may provide this benefit.

The results of Experiments 3-5 can be taken as support for the possible use of an “external memory” to help increase enumeration accuracy. Observers showed enhanced performance when they kept a “tally” of their count even in a spatially-irrelevant manner (Experiments 3A and 3B). This enhancement was not due to simply using a nonverbal report. As seen in Experiment 4, a nonverbal clicking response without a visible tally produced more errors than the pointing or tally methods, but still showed better performance than the cardinal report method. Locations, however, do play an important role in this task. Experiment 5 showed that even visiting the object locations and clicking at that location (with no persisting tally for visual aid) allowed for a higher enumeration range that was nearly identical to the condition where the markers remained on the screen (with a minor advantage for the standard pointing method in the 50-ms condition). There may be something about assigning a cardinal label to a set of items in working memory that produces more interference than a simple tally using external markers.
Another possible strategy that may be present in the current study is one that was discussed in Trick (2008). As previously described, observers can use a “group and add” strategy when enumerating displays of 1-9 items that were comprised of 1-3 shapes. The distinct shapes encouraged a grouping based on similarity and facilitated how quickly observers responded while increasing their subitizing capacity. This performance was explained by the visual indexing mechanism, which can rapidly select a subset of objects. Each subset’s numerosity was transferred into working memory in order to compute the overall numerosity by adding these subgroups. In the current study, the “group and add” strategy may be implemented according to proximity or the statistical regularity between objects. Regular spacing can produce canonical dot patterns that are easier to group and enumerate. For example, canonical displays with nine items tend to be grouped into triplets (Mandler & Shebo, 1982). This indicates a strategy of grouping into three-item sets and adding these “chunks” to obtain numerosity. Since grouping occurs in 80-120 ms (Kurylo, 1997), the 50-ms presentations may have prevented grouping and thus produced a lower subitizing range. The increased subitizing range observed for longer presentation durations may result from the ability to group items for more efficient processing. This explanation, however, still does not account for the observed disadvantage in the cardinal reporting method. Nevertheless, examining the response data and determining if observers display any sort of grouping strategies warrants further study.

Localization performance

Performance in reporting locations was highly accurate (average error distance $2.5^\circ$), compared to the mean distance between stimulus objects (over $6^\circ$). Location accuracy, however, decreased as the number of objects increased for displays with 2-6
objects, and this pattern held even when looking at errors for the first-click responses (Figures 30-33). These results follow those previously reported in experiments that examined the effect of memory load on object location accuracy, where localization errors increased with the number of items that had to be remembered (e.g., Dent & Smyth, 2006; Postma & De Haan, 1996). Since the current study indicates that enumeration performance only decreased when there were more than five or six objects (Figures 28 and 29), observers’ enumeration performance seems to be 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). Observers were able to individuate and report objects even when their report of locations was relatively impaired (i.e., in the 2-6 range). This finding is compatible with claims in Pylyshyn (2003, 2007) that encoding location occurs both temporally and logically after individuation. Moreover since encoding location may involve focal attention, it may be disrupted by a secondary task that demands attention, such as assigning a cardinal label.

The current results also provide some evidence on how global attention affects localization and enumeration. Previously reported results indicate that subitizing can occur outside of focal attention (Trick & Pylyshyn, 1994a). This is supported by the current results on 50-ms displays where verbal counting was not possible and serial attention was restricted. Increasing the display duration to 200 ms, which allows minimal serial attention, improves localization performance for displays with 2-6 items but improves enumeration performance only for displays with six or more items. Increasing exposure from 50-ms to 200-ms allows attentional processing that improves the encoding of location information. Visual indexes can serve this purpose by laying the groundwork
for object file representations, which hold more detailed information about object features. The briefest conditions, however, still produced a general localization accuracy that maintained the shape of the disc arrays. This may be the result of “visual routines” that execute scene parsing by extracting spatial relations and various properties from the raw input of early visual processing, which can assist in marking locations (Ullman, 1984). Since the pointing task requires a memory for locations, the information from visual routines can be used to aid performance when pointing to locations. These results also support the “coarse-to-finer” account for visual perception (Amit, et al., 2004).

A strong bias that occurred in the observers’ localization responses was a tendency to compress locations and report objects as being closer to each other (Figures 39 and 40). This may be due to errors in making the pointing movements with the mouse, but spatial compression also has been shown by studies on spatial representations in visual working memory (e.g., Ross, Morrone, & Burr, 1997; Sheth & Shimojo, 2001). In the current study, compression occurs around the centroid of the stimulus displays (as opposed to the central fixation) and localization responses seem to be “anchored” around this centroid. Such localization errors also could be attributed to the motor responses made in these tasks or to the effects of stimulus appearance in relation to the fovea, where foveated items have the benefit of a higher attentional resolution (He, Cavanagh, & Intriligator, 1996). In support of this account, stronger compression effects were observed when discs appeared further away from the point of fixation and into the periphery (Figure 35). Location on the fovea did not affect enumeration performance, since the displays were designed to encourage discriminability and prevent crowding even when items appeared in the periphery.
The effect of both sensitivity to regularities in the display and production of regularities in the responses are compatible with the limited bandwidth of the perception-memory channel (Miller, 1956). Object individuation is influenced by the spatial relationships among visual stimuli, including geometric factors such as symmetry, good continuation, and parallelism (Feldman, 2007). Additionally, several aspects of perception, including the individuation process, have processing qualities that can be described through information theoretic accounts (Brady, et al., 2009; Broadbent, 1965; Taylor, Maddess, & Nagai, 2008). A related result from the pointing task indicates that displays with less complex spatial relationships (i.e., more regularity) were processed more efficiently and produced better enumeration and localization performance (Figures 41 and 42).

The spatial representations of object location appears to be susceptible to systematic encoding errors and sensitive to the statistical regularity on a display. A recent study by Brady & Alvarez (2011) showed that the location of individual objects in visual working memory are not represented independently but depend on the relationships between objects in the form of “ensemble statistics”. This kind of summary information can bias the memory for the individual objects, such as their size (Brady & Alvarez, 2011). Additionally, ensemble statistics can be used to increase the amount of information that can be retained (Alvarez, in press; Alvarez & Oliva, 2009). In the current study, better localization performance on displays with higher regularity may result from the more compact information that was encoded for these displays (e.g., in a summary representation). Modeling the current data using these principles will be an
important next step to better understand the mechanisms underlying enumeration and localization.

Issues to consider

The data from these experiments indicate a fair amount of between-subject variability. This is especially evident in the shifts of the “knee” in performance (where accuracy drops): the subitizing range for individual observers ranged from 2-4 items to 2-8 items; however, most were near the 2-6 range. Additionally, localization performance varied among subjects, where some showed no effect of duration on accuracy. To improve the interpretation of these data, a version of the pointing task should be administered to a few observers over multiple sessions to provide more stable response data (e.g., administer 60 trials per condition instead of only 10). This should give a better idea about the effects of numerosity, duration, and reporting method, especially if the same trends can be identified reliably among different participants.

Another possible issue is the effectiveness of the mask used in the current study. Mandler & Shebo (1982) used 100, 200, 400, and 800-ms stimulus presentations, followed by a random-dot mask that appeared until a response was made in some experiments, or appeared after a response was made in other experiments. Similar to the results from the current study, Mandler & Shebo found that 200-ms displays produced more stable data than 100-ms displays. The stimuli in the current study were designed to be masked easily (e.g., low contrast and luminosity), but the duration of the 85-ms mask may have been more effective on the 50-ms displays than the longer durations. Some recommend using a mask that is longer than the stimulus (e.g., Sperling, 1960), but a decision was made to keep the mask constant throughout the current study in order to...
maintain a constant interval between the stimulus offset and the response screen. To explore the effect of mask duration on the pointing task, a follow-up experiment was conducted on 16 participants and used a 350-ms mask (but otherwise was identical to Experiment 1). This experiment produced comparable performance in the 200-ms condition (2-6 range for subitizing), but a slight decrease was observed for 50-ms and 350-ms displays (2-5 range). This may be attributable to memory decay since there was a longer duration between stimulus offset and the response screen, but further experiments testing this interpretation are warranted.

Finally, additional experiments that compare the pointing task to a cardinal reporting methodology should be conducted. For example, obtaining a cardinal response via a voice recorder or numbered response keys can be implemented in order to confirm the advantage observed in the pointing response method. The subitizing range observed in the numeral selection task, however, does not deviate from those reported in previous subitizing studies, and therefore may not be a concern. Nevertheless, additional iterations of this crucial comparison should be implemented since the results were unexpected and still require explanation.
Chapter 6. Conclusions

The current study examined the role of spatial information when enumerating small sets of objects. Additionally, the differences between various reporting methods in this enumeration task were examined. In particular, comparisons between the pointing task that involved a spatio-motor response was compared to the usual verbal count. Both the numeral selection task and the pointing task indicate how many items had been attended, but the pointing task revealed this number without requiring an explicit counting task. There was some reason to think that such a response, which may use the dorsal visual stream, may show a different pattern than the symbolic numeral response.

Although the reported findings require further empirical support, they do provide unique information about the individuation process and its interaction with the way in which observers report numerosity. The spatial configurations of the discs on the stimulus displays were retained relatively well, even when presented for only 50-ms. These representations, however, were susceptible to systematic localization errors, where observers compressed the distances between the objects. These errors increased for larger numerosity displays and were more pronounced in the shortest display duration. This suggests that location accuracy depends on further processing past the individuation stage. The individuation stage also was sensitive to scene statistics: regularity in object spacing increased enumeration and localization performance. Enumeration accuracy, however, was not impacted significantly within the subitizing range and benefited from increased exposure durations only for displays with more than six items.

The observed enumeration performance in the various tasks suggests that people may “see” more than they report when the reporting method is verbal in nature (i.e., by
providing a cardinal label). When nonverbally reporting this representation, for example by generating a number of clicks in the tallying task, performance improves. Performance is further enhanced when the tally can be placed on the locations of the stimulus discs, as in the pointing task. There are several candidate hypotheses for how the pointing method might help to increase the span of recall or “enumerating” in this subitizing experiment, such as the use of motor memory resources, an external memory, or an indexing mechanism. These provide theoretical challenges as well as ideas for further experiments that may support one or another of these options.

An important issue to address is the manner in which subitizing is defined across studies. Essentially, this asks at what stage does the subitizing limit occur? Is it during the individuation stage, where a number of objects are selected for processing? Or must the numerosity be serially identified or labeled in some way (e.g., by counting external objects or the number of active indexes)? If one takes the individuation of distinct visual items as sufficient for subitizing, then the limit in this study is up to six items. If a verbal report of numerosity of the individuated objects is necessary for subitizing, then the limit is around four items. The approach favored in this study is that the capacity for subitizing is the capacity for the individuation and selection of multiple objects. The manner in which observers report numerosity affects the upper limit, with a clear disadvantage for the symbolic report. Therefore, the non-symbolic reporting methods appear to provide a better reflection of the true capacity for the individuation of objects for enumeration, whether the reporting method is “tallying” a response or pointing to object locations. This perspective emphasizes the role of the indexing mechanism in small-set enumeration that works in addition to or instead of the magnitude estimation mechanism. This two-
mechanism view is feasible since the visual indexing mechanism already provides an account for the individuation and selection of items for perceptual processing, including enumeration. This has been supported by various studies and continues to be addressed in recent work (e.g., Chesney & Haladjian, under review; Haladjian, Chesney, & Pylyshyn, 2010).

The ability to individuate and enumerate sets of objects certainly warrants continued study. It would be especially illuminating to examine the information processing component of enumeration using this “pointing” method. For example, much can be learned from the strategies that are implemented to increase processing capacity, such as grouping or “chunking” strategies that have been shown to be used to increase localization accuracy (e.g., Alvarez, in press; Pothos & Chater, 2002; Sargent, Dopkins, Philbeck, & Chichka, 2010). In the current study, it would be useful to identify whether or not the stimulus discs are grouped by proximity and stored in working memory as compressed “chunks” (e.g., as shapes), which may be indicated by the spatial relationships among items. The current results support this possibility since displays with more spatial regularity show enhanced enumeration and localization performance, perhaps because certain groupings require less information to be encoded. The strategies of “chunking” groups of dots together and grouping by proximity will be explored in subsequent analyses of the localization data (in collaboration with Dr. Fabien Mathy, see Mathy & Feldman, 2009; Mathy & Feldman, under review); this project will examine whether the order in which the observers made their responses exhibit some sort of strategy that enhances performance.
Understanding how spatial information affects enumeration performance is an important project to undertake—one that has not been studied extensively. While manipulations of attention have shown that subitizing can occur without focal attention, exactly how spatial characteristics affect the ability to enumerate and localize objects requires further examination. The aim of the current study was to explore the relationship between enumeration and localization. The experiments described are a preliminary exploration of the results from the enumerating-by-pointing method, which indicates a higher enumeration capacity than other reporting methods and provides rich information about how observers perceive and represent sets of objects for enumeration. Further analyses of these results, along with additional experimental manipulations of the pointing task, will continue to provide insights on the processes involved in small-set enumeration.
Appendix

Figure A. Proportion of trials enumerated correctly in Experiment 2A, split by participant (display durations combined).
Figure B. Proportion of trials enumerated correctly in Experiment 2B, split by participant (display durations combined).
Figure C. Proportion of trials enumerated correctly in Experiment 3A, split by participant (display durations combined).
Figure D. Proportion of trials enumerated correctly in Experiment 3B, split by participant (display durations combined).
Figure E. Proportion of trials enumerated correctly in Experiment 4, split by participant (display durations combined).
Figure F. Proportion of trials enumerated correctly in Experiment 5, split by participant (display durations combined).
Figure G. Average miscounts (absolute value) with a single line for each subject (n = 152; n = 24 in 350-ms duration).
Figure H. The cumulative distribution for enumeration accuracy in 50-ms & 200-ms displays (n = 152). The x-axis is logarithmic so that equal intervals on the x-axis represent equal multiplicative deviations from the correct number. The fraction correct is given by the length of the solid vertical that superposes on the dashed vertical of the same color.
Bibliography


Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences, 102*(39), 14116-14121.


Curriculum Vita

Harry Haroutioun Haladjian

**Education**

2005 – 2011  Rutgers University, New Brunswick, NJ  
Ph.D., Cognitive Psychology

2005 – 2008  Rutgers University, New Brunswick, NJ  
M.S., Cognitive Psychology

1991 – 1995  Boston College, Chestnut Hill, MA  
B.A., Philosophy and Psychology

**Work Experience**

2010 - 2011  Teaching Assistant (Expository Writing), English Department, Rutgers University, New Brunswick, NJ

2006 - 2008  Teaching Assistant (Cognition Lab), Psychology Department, Rutgers University, New Brunswick, NJ


2001 - 2006  Research Consultant, Public Health Institute, Oakland, CA

2004 - 2005  Research Intern, Gabrieli Cognitive Neuroscience Laboratory, Stanford University, Stanford, CA

2001 - 2005  Computer Research Associate II, Prevention Research Center, Stanford University School of Medicine, Stanford, CA

2001 - 2005  Social Science Research Assistant III, Prevention Research Center, Stanford University School of Medicine, Stanford, CA

1997 - 2001  Social Science Research Assistant II, Prevention Research Center, Stanford University School of Medicine, Stanford, CA

1996 - 1997  Research Assistant and Bookkeeper Assistant, Jones Graduate School of Administration, Rice University, Houston, TX


