Vision Research 107 (2015) 133–145

Contents lists available at ScienceDirect

Vision Research

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

A snapshot is all it takes to encode object locations into spatial memory

Harry H. Haladjian[a,*], Fabien Mathy[b,1]

[a] School of Social Sciences and Psychology, University of Western Sydney, Australia
[b] Département de Psychologie, Université Nice Sophia Antipolis, France

ARTICLE INFO

Article history:
Received 1 May 2014
Received in revised form 17 December 2014
Available online 31 December 2014

Keywords:
Spatial attention
Visual short-term memory
Short-term memory
Subitizing
Grouping
Clustering analysis

ABSTRACT

This study examines the encoding of multiple object locations into spatial memory by comparing localization accuracy for stimuli presented at different exposure durations. Participants in the longest duration condition viewed masked displays containing 1–10 discs for 1–10 s (durations typically used in simple span tasks), and then reported the locations of these discs on a blank screen. Compared to conditions that presented the same stimuli briefly for 50 or 200 ms (exposures more typical of simultaneous spatial arrays), localization accuracy did not improve significantly under longer viewing durations. Additionally, a clustering analysis found that responses were spread among different clusters of discs and not focused on individual clusters, regardless of viewing duration. A second experiment tested this performance for displays containing two distinct clusters of discs to determine if clearly grouped subsets of objects would improve performance, but there was no substantial improvement for these two-cluster displays when compared to displays with one cluster. Overall, the results indicate that spatial information for a set of objects is extracted globally and quickly, with little benefit from extended encoding durations that should have favored some deliberative form of grouping. Such results cast doubt on the validity of Corsi blocks or equivalent common neuropsychological tests purportedly designed to evaluate specifically spatial short-term memory spans.

1. Introduction

Visual memory is often studied to identify the stages in perception with information processing limitations. The capacity limit in memory, for example, can affect the quality of simultaneous object representations, where having to remember more objects reduces the amount of detail that can be encoded about those objects (e.g., Alvarez & Cavanagh, 2004; Ma, Husain, & Bays, 2014). Others argue that there is a limit on the total number of objects remembered regardless of the amount of information encoded per object—the often cited “four slots” limit found in working memory studies (e.g., Cowan, 2001; Luck & Vogel, 1997; Zhang & Luck, 2008). Such processing limits also relate to the fast and error-free counting of up to four items, called “subitizing”, where enumeration errors and response latencies increase substantially for sets larger than four (e.g., Burr, Turi, & Anobile, 2010; Dehaene & Cohen, 1994; Kaufman et al., 1949; Pylyshyn, 1989; Revkin et al., 2008; Trick & Pylyshyn, 1993, 1994). Thus, it may be possible that both visual memory for objects and enumeration share a common resource with similar capacity limitations and variations that correlate between subjects (Cutini & Bonato, 2012; Piazza et al., 2011). This has been hypothesized to result from an initial competitive process of the individuation of the objects present in a visual scene (Melcher & Piazza, 2011).

The number of items that can be processed quickly (i.e., the subitizing range), however, tends to vary depending on the stimuli and reporting methods used. For example, there is an interaction between the intensity of the stimulus and the duration of exposure to it, with higher intensity stimuli requiring less time for detection (Hunter & Sigler, 1940) while also facilitating the detection of larger sets of items (Palomares & Egeth, 2010). Typical verbal reports produce a subitizing range of around four items (e.g., Revkin et al., 2008; Trick & Pylyshyn, 1993, 1994), with higher ranges when polygons forming prototypical configurations of dots are used, like the patterns found on a dice (e.g., Mandler & Shebo, 1982; Yantis, 1992). A recent localization study also identified a higher subitizing range when participants reported numerosity by marking the locations of briefly-viewed objects (Haladjian & Pylyshyn, 2011). In that study, participants were shown masked displays with randomly-placed discs at brief durations (50–350 ms), and then

---

* Corresponding author at: School of Social Sciences and Psychology, University of Western Sydney, Bankstown Campus, Building 24, Locked Bag 1797, Penrith, NSW 2751, Australia. Fax: +61 02 9772 6757.

E-mail addresses: haroutioun@gmail.com (H.H. Haladjian), fabien.mathy@unice.fr (F. Mathy).

[1] Département de Psychologie, Laboratoire BCL: Bases, Corpus, Langage – UMR 7320, Université Nice Sophia Antipolis, Campus SJA3, 24 avenue des diables bleus, 06357 Nice Cedex 4, France.

http://dx.doi.org/10.1016/j.visres.2014.12.014

marked the locations of each disc on a blank computer screen. In addition to measuring spatial memory for sets of objects, this reporting method provided a numerosity estimate. Enumeration performance was high for displays with up to six items when using the localization method, but only up to four items (the "typical" subitizing limit) when using a conventional reporting method with Arabic numerals in that study.

The motivation for the current study is to better understand the factors that may enhance spatial memory for a number of simple objects (e.g., Franconeri, Alvarez, & Enns, 2007; Haladjian & Pylyshyn, 2011) when using this location-based reporting method. One explanation for a higher capacity for remembering object locations may be related to the act of "pointing" to the locations of the discs, since this also engages a memory involved in motor responses (e.g., Goodale & Milner, 2004). Another possible explanation for this increased capacity is perceptual grouping, where nearby discs are grouped together for more efficient storage (e.g., Anderson, Vogel, & Awh, 2013; Brady & Tenenbaum, 2013; Feldman, 1999; Korjoukov et al., 2012). Effectively, a grouping process involves the ability for proximal discs to form a group and produce non-independent spatial information for those discs, which could be encoded compactly into a single "slot" in memory. This would allow the encoding of information about other discs (or groups of discs) into the remaining free "slots", and thereby increase the number of individual items that can be encoded. Such abilities for information processing systems to overcome capacity limitations whenever relational information can be computed has received particular attention recently in the visual short-term memory literature (e.g., Alvarez & Cavanagh, 2004; Bays, Catalao, & Husain, 2009; Bays, Wu, & Husain, 2011; Brady, Konkle, & Alvarez, 2009, 2011; Fougnie & Alvarez, 2011; Sargent et al., 2010; Wheeler & Treisman, 2002; Xu, 2002). The present study tests this possible explanation within a localization task by enhancing grouping effects with longer viewing durations (Experiment 1) and by presenting displays that have clearly groupable sets of objects (Experiment 2).

The manner in which object locations are encoded into memory can be described in two different ways. One view proposes that resource allocation is continuous (e.g., Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Gorgoraptis et al., 2011; Ma, Husain, & Bays, 2014; Wilken & Ma, 2004), which suggests that the key factor for memorization is the accuracy with which all the material is encoded. A contrasting view is that resource allocation is discrete, or slot-based, which proposes that the key factor for memorization is the number of objects that can be encoded (e.g., Donkin et al., 2013; Luck & Vogel, 1997; Zhang & Luck, 2008).

Resource allocation also can be framed in terms of information compression (e.g., Brady, Konkle, & Alvarez, 2009). One information compression method encodes information in a "lossless" manner (e.g., Mathy & Feldman, 2012), which allows the exact original data to be reconstructed from memory. In terms of spatial memory, it is possible that exact information about groups of items can be compressed in a lossless manner so that a greater number of items can be unpacked from a few groups. Similar to the ability to recall a series of 50 numbers, such as 2-4-6-8-10-, . . ., 100, by retaining the shorter description "even numbers from 2 to 100", it might be possible to retain the coarse locations of several groups of items (e.g., Aksentijević, Elliott, & Barber, 2001; De Lillo, 2004; Dry, Preiss, & Wagemans, 2012; Feldman, 1999; Korjoukov et al., 2012) without the loss of the original information regarding the number of items within each group. This process that supports the encoding of local perceptual structures, however, does not prevent any subsequent forms of distortion of the represented structures within groups. If present, this preliminary encoding of local structures can be detected using specific analyses. For example, this lossless encoding of local groups would produce a correct report of a limited number of items, with total loss of information for items that could not be encoded due to capacity limitations (essentially indicative of a slot-based fixed resource). One example is when an observer is shown a display with seven items, and she could encode a group of three items on the top of the screen and another group of two items on the left bottom part of the screen. This observer would in this case perfectly report the presence of two groups, and would report five discs with great accuracy, but would not correctly report the presence of the two other remaining discs on the right bottom part of the screen (again, this approach does not expect the individual locations to be reported perfectly for any of the discs).

An alternative encoding method that may help increase capacity can be described as "lossy". This may include the computing of a summary statistic, such as a global summary of spatial relationships (e.g., Jiang, Olson, & Chun, 2000; Sargent et al., 2010). Reproducing this information will result in more systematic errors distributed among all objects in memory and would be indicative of a more continuous and flexible resource model (e.g., see Franconeri, Alvarez, & Cavanagh, 2013). This would suggest a non-independent encoding of spatial information. (By analogy, these two forms of compression are similar to digital file formats such as .png/.gzip or .jpeg/.mpeg, which are respectively lossless and lossy.)

In the current study, we examine whether or not perceptual grouping of proximal objects improves spatial memory and capacity, and also characterize how spatial information tends to be encoded (i.e., lossless or lossy). Clustering measures were used to determine if participants use grouping to remember the number of items and possibly encode more precise spatial information about multiple object locations. A first hypothesis is that displays with more groupable arrays generally will facilitate spatial memory by inducing the grouping of a limited set of proximal objects, thus increasing capacity. A second hypothesis is that longer viewing durations enhance grouping by enabling more deliberate grouping processes. In this case, more objects are encoded more precisely due to the perception of a local structure since the focus of attention can be directed on individual groups. Alternatively, spatial information may be encoded globally in a quick "snapshot" and if so, would not benefit from longer exposures. In this latter case, object locations are encoded via the perception of a global structure (i.e., from a more diffuse or global focus of attention).

To investigate these questions, we used the localization task described above that required participants to remember object locations on displays containing randomly-placed discs, and we compared performance between exposure durations typical of simple span tasks (Shipstead, Redick, & Engle, 2012, p. 629) and change detection or continuous report tasks (Brady, Konkle, & Alvarez, 2011, p. 3). This essentially corresponds to comparing performance when local encoding is encouraged, to performance when global encoding is likely to occur. Note that we use a "pseudo-opposition" between the short-term memory durations only to reflect the fact that, generally, simple span tasks use longer presentation rates (to aid neuropsychological assessment and to facilitate instructions and computation of memory span) while rapid displays prevent the use of various conscious strategies. In typical simple short-term memory span tasks, the stimulus (verbal or spatial) is usually presented at a rate of one item per second (e.g., Gmeindl, Walsh, & Courtney, 2011 for the Corsi block-tapping test; see also the computerized spatial short-term memory tests of Lewandowsky et al. (2010)); the present study examines the gain that is expected with such longer durations. To find a common procedure for reporting the locations in the present study, however, both our conditions used a free recall procedure of the whole display in order to focus on grouping processes, rather than using either single-probe or whole-display recognition (see Rouder