Acta Psychologica 169 (2016) 100–108

Contents lists available at ScienceDirect

Acta Psychologica

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

Gist in time: Scene semantics and structure enhance recall of searched objects

Emilie L. Josephs [a,*,1], Dejan Draschkow [d,1], Jeremy M. Wolfe [b,c], Melissa L.-H. Võ [d]

[a] Cognitive and Neural Organization Lab, Harvard University, Cambridge, MA, USA
[b] Visual Attention Lab, Brigham and Women's Hospital, Boston, MA, USA
[c] Harvard Medical School, Cambridge, MA, USA
[d] Scene Grammar Lab, Johann Wolfgang Goethe-Universität, Frankfurt, Germany

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form 24 February 2016
Accepted 20 May 2016
Available online 3 June 2016

Keywords:
Repeated search
Scene perception
Object recall
Scene gist
Integration time
Task effects
Incidental memory
Semantic guidance

ABSTRACT

Previous work has shown that recall of objects that are incidentally encountered as targets in visual search is better than recall of objects that have been intentionally memorized (Draschkow, Wolfe, & Võ, 2014). However, this counter-intuitive result is not seen when these tasks are performed with non-scene stimuli. The goal of the current paper is to determine what features of search in a scene contribute to higher recall rates when compared to a memorization task. In each of four experiments, we compare the free recall rate for target objects following a search to the rate following a memorization task. Across the experiments, the stimuli include progressively more scene-related information. Experiment 1 provides the spatial relations between objects. Experiment 2 adds relative size and depth of objects. Experiments 3 and 4 include scene layout and semantic information. We find that search leads to better recall than explicit memorization in cases where scene layout and semantic information are present, as long as the participant has ample time (2500 ms) to integrate this information with knowledge about the target object (Exp. 4). These results suggest that the integration of scene and target information not only leads to more efficient search, but can also contribute to stronger memory representations than intentional memorization.

1. Introduction

Suppose that you wanted to learn what objects were present in a room with the goal of being able to recall those objects later. One way to do this would be to examine the room, intentionally trying to memorize the objects. However, explicit memorization is not the only way to encode information about objects encountered in the world. We also acquire memory representations for objects incidentally when, for example, we search for an object without an explicit instruction to memorize (Castelhano & Henderson, 2005; Draschkow & Võ, 2016; Hout & Goldinger, 2010; Hout & Goldinger, 2012; Howard, Pharaon, Körner, Smith, & Gilchrist, 2011; Olejarczyk, Luke, & Henderson, 2014; Võ, Schneider, & Matthias, 2008).

Are intentionally memorized and incidentally encountered objects encoded differently? This seems likely, given that different tasks (e.g. search and memorization) require different interactions with the same stimuli. For instance, Rothkopf, Ballard, and Hayhoe (2007) showed task-specific deployment of attention to different parts of the same scene or objects. They found that the proportion of fixations landing on obstacles compared to targets changes depending on whether the observer is told to avoid the obstacles or collect the targets. Additionally, the authors demonstrated that different areas of the objects are selected for fixation in each of these tasks. Task-specific effects on eye movements are also evident in saccade lengths, which are longer in search tasks than in free viewing, and in the local image characteristics of the areas selected for fixation (Tatler, Baddeley, & Vincent, 2006). In addition to such differences in the deployment of eye movements, task-specific cognitive requirements also seem to cause differences in the extraction of information from a single fixation. For example, Tatler and Tatler (2013) found that task irrelevant objects received the same number of fixations when participants were told to memorize all the objects in the scene as when they were told to memorize a specific subset of objects (e.g. only objects used to make tea), yet object memory was higher in the first case.

Do these task-dependent modulations of attention and information extraction produce differences in recall between objects that have been memorized and objects that were searched? In the work of Võ and Wolfe (2012), participants searched for objects in scenes. Participants located targets in scenes more quickly if they had searched for them in a previous block. They showed no such improvement if they had been familiarized with the scene in other ways, such as searching

* Corresponding author at: Cognitive and Neural Organization Lab, William James Hall, 720, 33 Kirkland St, Cambridge, MA 02138, USA.
E-mail address: ejosephs@g.harvard.edu (E.L. Josephs).
[1] authors contributed equally.

http://dx.doi.org/10.1016/j.actpsy.2016.05.013

for letters superimposed on the targets, exploring the scene for 30 s to determine if a man or woman decorated the room or by explicitly trying to memorize the scene prior to the search. A follow-up by Hollingworth (2012) also demonstrated that a previous search for a target speeds a subsequent search more than memorizing object locations or evaluating semantic properties of the scene. It seems that while looking *at* an object in a scene certainly creates a memory trace for it, looking *for* an object can build memory representations that can be used to find the same object again more efficiently (Võ & Wolfe, 2012; for a review see Võ & Wolfe, 2015).

These studies suggested differences between the representations formed incidentally during search and those formed intentionally during memorization. However, they did not measure recall explicitly, instead inferring target memorization from improved reaction times. In order to more directly compare representations created during search with those created during memory, Draschkow et al. (2014) used a free recall task. In their study, participants performed one block where they searched for objects in photographs of scenes, and another where they memorized objects in a different set of scenes. Each block was followed by a free recall test in which participants were asked to draw all the objects they could remember. A comparison of the average number of drawn targets revealed better recall following the search block than following the memorization block.

The first two experiments in the Draschkow et al. (2014) paper showed a substantial effect of task on recall rates of target objects. However, these results left open the question of whether all search tasks, regardless of stimulus set, lead to stronger memory representations, or whether this effect is only present when search is being performed in a naturalistic scene. Searches through scenes rather than displays of isolated items have been shown to make use of a very rich array of semantic scene information (for a review see Henderson, 2007; Wolfe, Võ, Evans, & Greene, 2011). Such recruitment of information may strengthen the representation of the searched objects over and beyond that of objects in non-scene contexts. In Experiment 3, Draschkow et al. (2014) tested the role of the scene content of the images by repeating their experiment using non-scene stimuli. They created images with unique textures as backgrounds (folds of fabric, droplets of water, a field of clover leaves), upon which they placed images of the objects that had been designated as targets in the previous experiments (see Fig. 1). The thumbnail images consisted of isolated exemplars of each of the original targets. These thumbnails were evenly distributed on the background images, removing any meaningful spatial relationships between the objects. Repeating the experiment with these non-scene stimuli abolished the original results: searched objects no longer showed a recall benefit over intentionally memorized objects.

These results indicated that simply searching for objects does not always build stronger representations than simply memorizing them. Performing the search in a meaningful, semantically rich scene seems to be important. However, the Draschkow et al. (2014) study could not specify why the effect was only observed in scenes. One possibility is that scenes are highly information-rich displays relative to randomly-organized collections of objects. In the process of transforming the scenes into non-scene stimuli in the Draschkow et al. (2014) study, objects were dissociated from their backgrounds, made uniform in size, and were placed at random locations on the screen. As will be described below, each of these sources of information has been shown to play a role in facilitating object perception or guiding search in scenes (Bar, 2004; Biederman, Mezzanotte, & Rabinowitz, 1982; Castelhano & Heaven, 2011; Torralba, Oliva, Castelhano, & Henderson, 2006). It is possible that accessing one or more of these sources of information during search for a target could create a stronger memory representation for that object than memorizing it.

Relationships Between Objects: One source of information exploited during real-world searches is the learned regularity in object grouping. Coffee mugs can be reliably found in proximity to coffee makers, pens are often found next to notebooks. Indeed, Castelhano and Heaven (2011) found that objects in their correct spatial grouping are easier to find and recognize than those in incongruent spatial groupings, even if the identity of the scene is made ambiguous by the presence of incongruent objects. Furthermore, ambiguous drawn object are more easily recognized if they are grouped with related objects (Bar, 2004). In general, objects in probable locations are better recognized than the

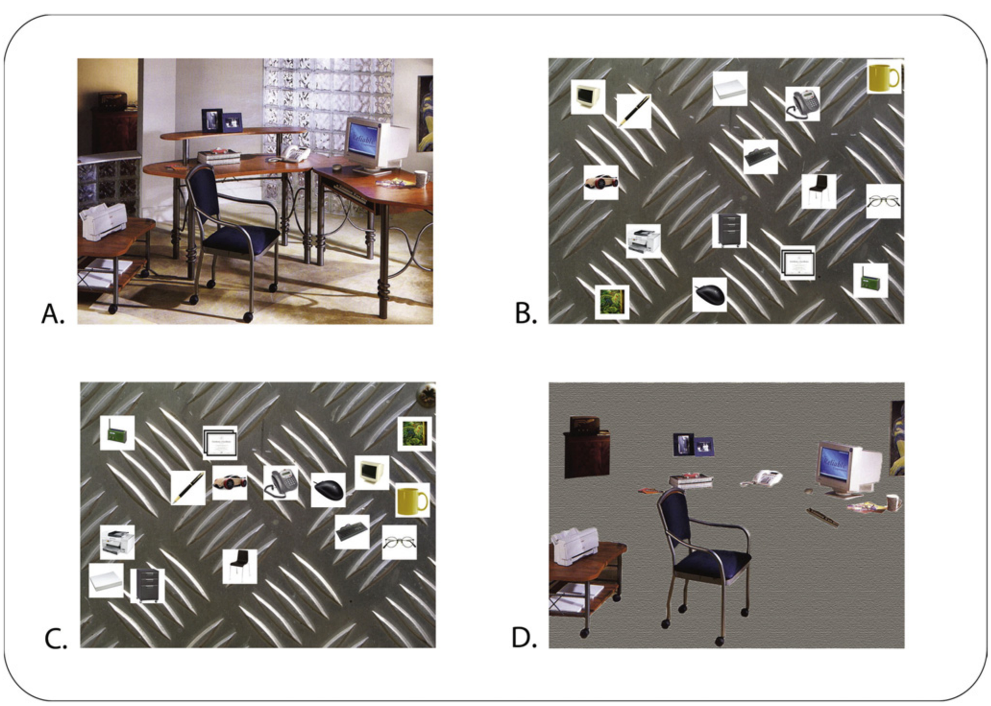

Fig. 1. A–B: Images used in Draschkow et al. (2014). B was created by finding new examplars of the targets selected in A and placing them on a non-scene background. C–D: images created for the current study. C was created by moving the objects from B into the spatial relationships they had in A. Notice how now the computer, keyboard and mouse are located near each other, as we would expect in real world situations. Stimulus D was created from stimulus A using image editing software to remove the background.