Consciousness and Cognition 30 (2014) 256–265

Contents lists available at ScienceDirect

Consciousness and Cognition

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

Iconic memory for the gist of natural scenes

Jason Clarke *, Arien Mack

The New School for Social Research, Visual Perception Lab, 80 Fifth Avenue, 7th floor, New York City, NY 10011, United States

ARTICLE INFO

Article history:
Received 11 August 2014

Keywords:
Iconic memory
Gist perception
Natural scene perception

ABSTRACT

Does iconic memory contain the gist of multiple scenes? Three experiments were conducted. In the first, four scenes from *different* basic-level categories were briefly presented in one of two conditions: a cue or a no-cue condition. The cue condition was designed to provide an index of the contents of iconic memory of the display. Subjects were more sensitive to scene gist in the cue condition than in the no-cue condition. In the second, the scenes came from the *same* basic-level category. We found no difference in sensitivity between the two conditions. In the third, six scenes from different basic level categories were presented in the visual periphery. Subjects were more sensitive to scene gist in the cue condition. These results suggest that scene gist is contained in iconic memory even in the visual periphery; however, iconic representations are not sufficiently detailed to distinguish between scenes coming from the same category.

1. Introduction

In a fleeting look at the world, we have the subjective impression of a richly detailed, panoramic scene. Glancing along Fifth Avenue, New York City, we feel that we momentarily register *all* of it: the busy sidewalk, the sky, vehicles, shapes, textures, colors, as well as an overall understanding of what we are looking at – the *gist* of the scene. While some scientists argue that this is due to a fleeting visual sensory memory of a rich phenomenal experience (Block, 2005, 2011; Koch & Tsuchiya, 2007; Lamme, 2010), others contend that we do not see as much of it as we think (De Gardelle, Sackur, & Kouider, 2009; Naccache & Dehaene, 2007; Noe, 2002; Rosenthal, 2007).

How much information do we see in a glance? When presented with a display containing items, subjects can typically report around 4 of the items in the display (Erdmann & Dodge, 1898; Luck & Vogel, 1997; Sligte, Scholte, & Lamme, 2008; Sperling, 1960). This visual short-term memory (VSTM) lasts on the order of seconds (Luck, 2007), and its apparently limited capacity has been used to support the claim that we see far less than we think we do, with the contents of our conscious experience being the contents of VSTM (Block, 2011; Lamme, 2010). However, many psychophysical experiments have demonstrated that observers have more information available about the contents of a briefly presented display than they can typically report. Using an ingenious procedure, the partial-report procedure, in which a cue required subjects to report only a sample of a display containing letters (partial report) instead of all of the items (whole report), Sperling (1960) discovered that indeed subjects are able to report nearly all of the items for about half a second or so (depending on the conditions) after offset of the display. This brief memory was dubbed *iconic memory* (Neisser, 1967), and many experiments over the last fifty years or so have discovered more about the processes and representations underlying it. The basic findings are that it is high capacity, short duration, and precategorical in nature (Dick, 1974; Sperling, 1960), that is it does

* Corresponding author.
E-mail address: clarj081@newschool.edu (J. Clarke).

http://dx.doi.org/10.1016/j.concog.2014.09.015
1053-8100/

not contain semantic or category information. However, there is conflicting evidence regarding the latter (e.g. Coltheart, 1980; Keysers, Xiao, Foldiak, & Perrett, 2005).

Do human observers have an iconic memory for the contents and the gist of a natural or real-world scene? As active perceivers, we are faced with the task of processing and understanding the everyday complex, naturalistic visual scenes before us. Natural scenes contain a seemingly infinite number of objects that can be categorized. Human observers are able to effortlessly recognize and categorize these objects despite variations in lighting, occlusion by other objects, and unusual viewpoints (Logothetis & Sheinberg, 1996). Furthermore, objects rarely (if ever) appear in isolation, but are seen as part of a meaningful context (the so-called *gist* of the scene) and in typical situations their identities are indeed constrained by the context, semantically and physically speaking: a table lamp is more likely to appear in a living room setting than in a forest, for example, and a fire hydrant is more likely to appear on the sidewalk than in the sky. Such regularities, invariants, and knowledge gained through experience might lead to efficiency in processing by the visual system. Indeed, many studies show that the brain may be particularly efficient at the processing of such naturalistic stimuli (Biederman, 1972; Biederman et al., 1973, 1974; Boyce & Pollatsek, 1992; Braun, 2003; Fei-Fei, Van Rullen, & Koch, 2002). The purpose of this research is to begin to answer the questions: What information is contained in iconic memory from briefly presented natural or real-world scenes? Does iconic memory contain information about the gist of multiple scenes?

Observers are remarkably efficient at extracting the gist or category from a glance at a natural scene when it is briefly presented (Potter, 1976; Fei-Fei, Iyer, Koch, & Perona, 2007; Fei-Fei et al., 2002; Schyns & Oliva, 1994), leading some researchers to claim that information about scene gist is preferentially processed in the brain (Fei-Fei et al., 2007). Indeed recent neurophysiological research provides evidence that a specific area of the brain, the parahippocampal place area (PPA), is involved in the processing of natural scenes. For example, using fMRI Epstein and Kanwisher (1998) found that the PPA responded selectively to passively viewed scenes but weakly to objects and not at all to faces. Moreover, they found that the response of the PPA was just as strong for empty rooms (and therefore just a spatial layout) as for rooms containing many objects, and the response disappeared when the spatial arrangement of the room was disrupted such that it no longer defined a coherent space (Epstein & Kanwisher, 1998). This suggests that natural scenes are a special and evolutionarily important kind of visual stimuli.

Psychophysical and neuroimaging studies have found that humans are remarkably efficient at categorizing scenes with "minimal attention" (Fei-Fei et al., 2002). Indeed, neuroimaging studies have shown that neural activity in cortical areas known to be involved in natural scene perception is present even without selective attention (Peelen, Fei-Fei, & Kastner, 2009). However, other studies have demonstrated the need for attention in consciously perceiving natural scenes (Cohen, Alvarez, & Nakayama, 2011; Mack & Clarke, 2012) as well as priming from natural scenes (Clarke, Ro, & Mack, 2013 Abstract). These later studies show that while gist is picked up easily, it still requires attention for it to be conscious and thus reportable.

Why should we be interested in whether the gist of natural scenes exists in iconic memory? For a few reasons at least: First, natural scenes are complex stimuli, and their presence in iconic memory supports a view of this memory as not only containing simple features but also high-level perceptual structure. Second, the gist of a scene (a beach, a bathroom etc) is information about the semantics or category of the display; therefore, its presence here would suggest that iconic memory is not precategorical but contains information about the meaning of the stimulus. Finally, demonstrating the presence of gist information in iconic memory suggests that neural areas, namely the PPA, that encode information about natural scenes, are active in iconic memory. As these areas are positioned higher up in the visual system, this would give further support for the view that iconic memory is a late process in the visual hierarchy (Keysers et al., 2005).

The procedure we used to assess the contents of the iconic memory for scenes was modeled on Sperling's partial versus whole report procedure which served him as the measure of the contents of iconic memory (e.g. Averbach & Coriell, 1961; Sperling, 1960). In all experiments, subjects were briefly presented with an array of scenes (for 250 ms in the first two experiments and 500 ms in the third experiment). In the no-cue condition, 200 ms after offset of the array, a one word gist descriptor e.g. "waterfall" appeared on the screen, and the subject's task was to report whether a scene fitting that description had been present. In the cue condition, immediately after offset of the array, a cue directing attention to one of the no-longer present scenes appeared for 200 ms and was followed by a one word gist descriptor. The 200 ms ISI in the no-cue condition is, therefore, replaced by the 200 ms cue in the cue condition. The subjects' task in the cue condition was to report whether the cued scene fit that description. While in many respects these measurements are similar to those used in earlier research, in the experiments reported here, subjects were not required to name the items present in the display (as they were in earlier studies), but instead to report whether a word following the display referred to any of the scenes in the no-cue condition or the cued scene in the cue condition. Following Sperling, our index of the contents of iconic memory was the difference between the number of items correctly reported in the cue and no-cue conditions (Sperling, 1960). We reasoned that while in the cue condition, the subject only has to inspect the cued scene (akin to the partial report condition for Sperling), in the no-cue condition the subject has to search more of the array: all of it when there is no matching scene and, on average, half of it when there is. While this is being done the memory is decaying. Thus, the no-cue condition may underestimate the information available in the iconic store. By the time the subject has searched through an internal representation of the array, it has disappeared. This search is not needed in the no-cue condition (akin to the partial report in Sperling's experiments), which, therefore, provides a more accurate index of the contents of iconic memory as it allows us to sample the information that is available immediately at display offset without requiring a search of rapidly decaying representations.