Neuropsychologia 47 (2009) 1994–2003

Contents lists available at ScienceDirect

Neuropsychologia

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

Fixation and saliency during search of natural scenes: The case of visual agnosia

Tom Foulsham[a,*], Jason J.S. Barton[b,c,d], Alan Kingstone[d], Richard Dewhurst[a], Geoffrey Underwood[a]

[a] *School of Psychology, University of Nottingham, UK*
[b] *Department of Medicine (Neurology), University of British Columbia, Canada*
[c] *Department of Ophthalmology and Vision Sciences, University of British Columbia, Canada*
[d] *Department of Psychology, University of British Columbia, Canada*

ARTICLE INFO

Article history:
Received 15 July 2008
Received in revised form 12 January 2009
Accepted 10 March 2009
Available online 18 March 2009

Keywords:
Agnosia
Eye movements
Scene perception
Visual search
Saliency

ABSTRACT

Models of eye movement control in natural scenes often distinguish between stimulus-driven processes (which guide the eyes to visually salient regions) and those based on task and object knowledge (which depend on expectations or identification of objects and scene gist). In the present investigation, the eye movements of a patient with visual agnosia were recorded while she searched for objects within photographs of natural scenes and compared to those made by students and age-matched controls. Agnosia is assumed to disrupt the top-down knowledge available in this task, and so may increase the reliance on bottom-up cues. The patient's deficit in object recognition was seen in poor search performance and inefficient scanning. The low-level saliency of target objects had an effect on responses in visual agnosia, and the most salient region in the scene was more likely to be fixated by the patient than by controls. An analysis of model-predicted saliency at fixation locations indicated a closer match between fixations and low-level saliency in agnosia than in controls. These findings are discussed in relation to saliency-map models and the balance between high and low-level factors in eye guidance.

1. Introduction

When humans look at a scene, they scan it with a series of fixations directed at different elements of the scene. The distribution of these fixations is not random, but at the same time they are not stereotyped: since the seminal studies of Yarbus (1967), it has been clear that there is great variability between observers in their scanpaths. Understanding the factors that guide the distributions of these fixations and generate this variability is important to our understanding of how natural scene perception is accomplished.

How is the scanning of scenes affected by deficits in object recognition? Visual agnosia is a neuropsychological impairment in which the patient is unable to recognize objects by sight, despite normal visual acuity and semantic knowledge (Farah, 1990; Humphreys & Riddoch, 1987; Riddoch & Humphreys, 1987). Time-honoured distinctions have been drawn between those whose agnosia can be attributed to problems in perceiving shape and form (apperceptive agnosics) and those whose perception is intact but disconnected from semantic associations (associative agnosics; Lissauer, 1890); in addition there are more recent descriptions of an integrative form, consisting of failure to combine elements into a coherent object, as typified by subject HJA (Behrmann, 2003; Riddoch & Humphreys, 1987).

The scanning of scenes containing natural objects by patients with visual agnosia is of interest given current distinctions being drawn between factors guiding fixation patterns, particularly between top-down and bottom-up guidance (Hayhoe & Ballard, 2005; Henderson, 2003). Top-down guidance refers to the influence on fixation patterns of a variety of observer- or task-dependent factors, such as task instructions (Dewhurst & Crundall, 2008), task goals (Foulsham & Underwood, 2007; Yarbus, 1967), prior knowledge (Althoff & Cohen, 1999), or perceptual expertise (Reingold, Charness, Pomplun, & Stampe, 2001; Underwood, Chapman, Brocklehurst, Underwood, & Crundall, 2003). In a search task, for example, the goal of the observer (to find an item) causes them to move their eyes towards targets or target-similar distractors (van Zoest, Donk, & Theeuwes, 2004; Wolfe, Butcher, Lee, & Hyle, 2003). Top-down search is therefore directed more to 'relevant' items in a scene. Bottom-up guidance, on the other hand, refers to the influence of stimulus properties or features such as colour, contrast and motion, properties that are not dependent on the observer or task: bottom-up search is directed towards conspicuous or 'salient' items.

Several models of eye movement control have been proposed to explain and predict the location and duration of fixations made during perception. Much of this work is found in the literature on reading and eye movements although there are related models which can also be applied to visual search and scene perception

* Corresponding author at: Department of Psychology, University of British Columbia, 2509–2136 West Mall, Vancouver, Canada V6T 1Z4.
E-mail address: tfoulsham@psych.ubc.ca (T. Foulsham).

0028-3932/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2009.03.013

(Engbert, Nuthmann, Richter, & Kliegl, 2005; Findlay & Walker, 1999; Logan, 1996; Reichle, Rayner, & Pollatsek, 2003). Findlay and Walker's (1999) framework specifies a system where two metrics are programmed for every saccade: WHEN the saccade should occur, and WHERE it should be directed. Normally this system regulates how long we look at objects, and the location our eyes select as the target of upcoming saccades to ensure optimal visual scanning. A key question is whether the operation of this system is affected by the defective object processing in visual agnosia. Could this lead to an extension of fixations durations or time spent dwelling upon objects which is comparable to increased activity in the WHEN pathway of Findlay and Walker's model?

The location of the next saccade within Findlay and Walker's (1999) framework is determined by the greatest saliency peak represented on a topographic map. In the past, quantifying this bottom-up saliency was difficult, although various statistics suggested that people were drawn to edges and areas of contrast (Mannan, Ruddock, & Wooding, 1996; Reinagel & Zador, 1999). Recent work has generated a well-specified model that combines visual features (colour, intensity and orientation contrast) at various spatial scales to produce an overall spatial representation of the bottom-up saliency of the different parts of a scene (Itti & Koch, 2000). The model assumes that attention and therefore fixations will be attracted towards the peaks of high saliency in this map, and thus predicts search patterns in which people move fixations and/or covert attention between scene elements in their descending order of saliency, predictions supported by data from human search times while looking for camouflaged military vehicles in outdoor scenes (Itti & Koch, 2000). The model also performs better than chance at predicting eye movements when scenes are free-viewed or inspected in preparation for a memory test (Foulsham & Underwood, 2008; Parkhurst, Law, & Niebur, 2002).

In other search tasks however, bottom-up saliency is not a good predictor of fixation distribution. Chen and Zelinsky (2006) found that when top-down and bottom-up signals were placed in competition, top-down guidance dominated, presumably by filtering out irrelevant items based on known target features and peripheral vision. The stimuli used in this study were greyscale, photorealistic objects arranged in a circular array, which removed real-life expectations about the probable locations of objects. Nevertheless, in more realistic search tasks, Foulsham and Underwood (2007) and Henderson, Brockmole, Castelhano, and Mack (2007) also found that the saliency-map model did not predict eye movements. In Foulsham and Underwood (2007) (see also Underwood & Foulsham, 2006; Underwood, Foulsham, van Loon, Humphreys, & Bloyce, 2006) participants viewed a set of natural photographs, some of which contained key objects which had been ranked by the saliency model, under several task conditions. When viewing the same scenes for a memory test, visual saliency had an effect on how early and how often these objects were inspected. However, when these objects became targets in visual search the effects of saliency were eliminated and targets were found quickly regardless of their saliency. Henderson et al. (2007) recorded the eye movements of observers searching for people in outdoor scenes and found that the saliency-map model did not predict these movements. As a consequence of such results Torralba, Oliva, Castelhano, and Henderson (2006) proposed a contextual guidance model of real-world search which biases eye movements with global scene features or scene "gist".

The implication of these studies is that for the healthy observer top-down knowledge "overrides" or weights a saliency-based guidance system, and that this information is available early, perhaps within the first fixation on a scene. This top-down knowledge might be in the form of expectations concerning target features or locations (as in models proposed by Navalpakkam & Itti, 2005; Rao, Zelinsky, Hayhoe, & Ballard, 2002; Zelinsky, Zhang, Yu, Chen, & Samaras, 2005), or it might consist of recognizing global scene properties and gist preattentively and using these to guide the eyes.

What happens when the ability to recognize object or global scene properties is absent or impaired? Under such circumstances it might be more difficult to implement top-down guidance. A person with visual agnosia might not be able to override a bottom-up saliency-based system, due to their inability to link raw visual input to top-down knowledge. If so, their fixation patterns should conform more closely to the predictions of the saliency-map model: that is, they should show a significant impact of the salience of stimulus properties on visual search in situations where normal subjects do not display such effects. Such a finding would extend the applicability of the saliency-map model to more naturalistic situations and support the tentative suggestion that such bottom-up effects are present but normally over-ridden by top-down considerations.

In this study we report on the fixation patterns made during visual search by a patient with general visual agnosia. Despite good visual acuity and peripheral fields, indicating functioning low-level vision, and apparently normal voluntary saccades, she is unable to recognize even simple three-dimensional abstract forms, or line drawings of common objects. How would such a patient distribute fixations in a scene filled with objects, particularly when given the instruction to search for a certain category of object? We hypothesized that, with her deficit, there should be less top-down guidance in this task than with normal controls. If visual saliency is computed earlier than, or independent from, object recognition, then saliency would be predicted to have an effect on eye movements. Furthermore, with reduced top-down bias, the eye movements produced might be closer to a raw saliency map than those made by normal controls.

2. Method

2.1. Case description

CH is a 63-year-old right-handed woman with slowly progressive visual difficulties over a period of 6 years. She first noted trouble with reading, especially large type, although she could still write. Subsequently she had difficulty recognizing the faces of her friends, relying on their voices instead. She had problems locating household objects, for example in the refrigerator or on the kitchen counter, and often misreached for items like light switches and cups. She confused navy with black but otherwise believed her colour vision to be normal. She later developed more problems navigating in familiar surroundings and required an escort on her visits to the clinic.

Her visual acuity was good (20/25 for single letters) and her visual fields were full, confirmed by Goldmann perimetry. Saccades and pursuit eye movements were normal. Neuropsychological evaluation showed normal general knowledge, expressive vocabulary and comprehension. She was able to write to dictation, with occasional spelling errors, but could not read words, proceeding instead by deciphering one letter at a time. Digit span was 6 forwards and 4 backwards, and she performed in the average range on all tests of auditory verbal learning and memory. Abstract verbal reasoning abilities were average to superior.

Visual tasks were severely impaired, with difficulty on line bisection, cancellation and search tasks. Her object recognition was slow and effortful, with severe impairments on the Boston Naming Test: with line drawings she frequently failed or misidentified objects, often focussing on small details (key = "see the O", chair = "grass. . .something to sit on"). She had difficulty naming three-dimensional shapes (cylinder = "cone", cube = "all these angles, octagon", pyramid = "triangles in it, a tent"), although she could name two-dimensional figures like squares and circles. She interpreted the Cookie theft picture as "a woman washing dishes, something spilling, it must be a restaurant". There was complete disorganization in copying or spontaneous drawing of a flower (Fig. 1).

Further perceptual tasks confirmed severe visual problems. Benton line orientation test showed 20% accuracy, in the severely deficient range, but curvature discrimination was normal. Her ability to judge the spatial configuration of dot patterns was impaired, scoring 56% correct with two dots and 22% correct with four dots (chance = 33% correct). Her ability to judge whether a triangle was symmetric or not was at chance (56% correct), a task controls do with 100% accuracy. Her ability to distinguish famous faces from anonymous ones was poor (d' = 0.12, versus d' for controls > 2.5). Imagery for famous faces was borderline for facial features (13/16) and impaired on overall face shape (11/16), with the lower limit of normal performance being 82% accuracy (Barton & Cherkasova, 2003).

A cerebral perfusion scan with technetium injection at 4 years after onset showed marked hypoperfusion of the posterior parietal and temporal lobes. CT