Available online at www.sciencedirect.com
ScienceDirect
Acta Psychologica 125 (2007) 129–143
acta psychologica
www.elsevier.com/locate/actpsy

Describing scenes hardly seen

Christian Dobel *, Heidi Gumnior, Jens Bölte, Pienie Zwitserlood

Psychologisches Institut II, Westfälische Wilhelms-Universität Münster, Fliednerstr. 21, 48149 Muenster, Germany

Received 14 December 2005; received in revised form 6 July 2006; accepted 7 July 2006
Available online 28 August 2006

Abstract

Knowledge about scene categories, the so-called gist, can be extracted very rapidly, while recognition and naming of individual scene objects is a more effortful process. We investigate this phenomenon by presenting action scenes involving two actors for durations varying between 100 and 300 ms. Incoherence was created by mirroring individual scene actors. Upon masked presentation participants had to report content, actors and objects and to indicate whether the scene was meaningful or not. Scene coherence was judged correctly at all presentation durations. Actors were correctly identified in about one-third of the cases even with presentation durations of 100 ms, and identification rate increased up to 80% with longer durations. Identification depended on scene coherence, on the position of agents in the scene, and on the position of actors relative to the fixation cross. These interdependencies of scene and object perception indicate that the visual system seems to be very sensitive to meaningful interactions of living entities. A series of fixations is not necessary to identify actors of a scene.

PsycINFO classification codes: 2323; 2346; 2720

Keywords: Scene apprehension; Language production; Event cognition; Object identification

1. Introduction

How can the gist of a complex scene be apprehended at a single glance if focused attention is crucial for the identification of individual objects? Our research addresses the issue of

* Corresponding author. Address: Institut für Biomagnetismus und Biosignalanalyse, Malmedyweg 15, 48149 Münster, Germany. Tel.: +49 251 8352430; fax: +49 251 8356874.
E-mail address: cdobel@uni-muenster.de (C. Dobel).

0001-6918/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2006.07.004

perceiving the whole and its parts. We focus on the fast uptake of information from complex scenes, involving two actors (the "parts") engaged in a meaningful or meaningless action (the "whole"). On the one hand, there is ample evidence that attention is crucial for conscious visual perception. The strongest position divides the mental world into a preattentive and an attentive world, with a crucial role for focused attention to bind features that are initially processed separately (e.g., Treisman & Gelade, 1980). On the other hand, it is known for decades that the "gist" of a scene can be recognized with very short presentation durations (Potter, 1975). These two positions seem mutually exclusive, so, what is the evidence for each?

In his overview on visual search, Wolfe (1998) describes the preattentive world as a level of representation inhabited by objects that can be searched for, but whose identity is not known: "Preattentive processes divide the scene into "things" and the preattentive basic features describe the "stuff" out of which perceptual "things" are made" (Wolfe, 1998, p. 43). This view is supported by studies employing visual search, but also by the phenomenon of change blindness (for overviews: Rensink, 2002; Simons & Rensink, 2005). In the "flicker" paradigm frequently used to study change detection (Rensink, O'Regan, & Clark, 1997), an original and a modified image continually alternate, with a brief blank field between the two. This produces a failure or delay in detecting changes between the two images. Such effects demonstrate that large visual changes can slip the viewer's awareness (even if the whole image moves; see Schofield, Bishop, & Allan, 2006). Fixating closely to or at the area of change improves detection in the flicker paradigm (Hollingworth, Schrock, & Henderson, 2001). Williams and Simons (2000) showed that change detection performance does not improve with training, and that changes are not easier to detect in moving than in static objects. Triesch, Ballard, Hayhoe, and Sullivan (2003) demonstrated that detection of object changes is task-dependent. Changes in object attributes can be detected when the attribute is task-relevant at the particular moment in time, otherwise changes are often missed. An obvious conclusion from these findings is that only those object features are perceived that are needed for the actual task at hand. Put differently, object perception is a task-dependent process, hardly influenced by training, or by visual object or change characteristics. Obviously, sequences of covert or overt attention shifts are necessary to identify objects in the visual field. Note however, that the fleeting character of the visual word does not preclude the presence of rich and detailed representations. It might be that they fade quickly, or do not reach systems responsible for change detection and/or overt responses (Simons & Rensink, 2005).

In stark contrast to the perception of specific objects and their attributes, some type of high-level representation of a visual scene is accessed within a fraction of a single fixation, i.e. within 30–50 ms (Biederman, 1972; Biederman, Rabinowitz, Glass, & Stacy, 1974; Hollingworth & Henderson, 1998; Potter & Levi, 1969; reviewed by Henderson & Ferreira, 2004; VanRullen & Koch, 2003). This representation is rich enough to apprehend the "gist" of a complex scene, that is, "knowledge of the scene category (e.g., kitchen) and the semantic information that may be retrieved based on that category" (Henderson & Ferreira, 2004, p. 15). Shifts of attention are not necessary to apprehend this basic scene-category information (Tversky & Hemenway, 1983). Recent demonstrations of the visual system's effectiveness for processing complex, natural scenes corroborate these findings. Thorpe and colleagues have repeatedly shown that highly detailed photographs can be accurately sorted into "living" and "non-living" categories, even with presentation durations as short as 20 ms (Delorme, Richard, & Fabre-Thorpe, 2000; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001).