Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 597–607
Vision Research
www.elsevier.com/locate/visres

# The duration of the attentional blink in natural scenes depends on stimulus category

Wolfgang Einhäuser [a,*], Christof Koch [a,b], Scott Makeig [c]

[a] Division of Biology, California Institute of Technology, Pasadena, CA, USA
[b] Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA, USA
[c] Swartz Center for Computational Neuroscience, University of California at San Diego, San Diego, CA, USA

Received 6 March 2006; received in revised form 26 November 2006

**Abstract**

Humans comprehend the "gist" of even a complex natural scene within a small fraction of a second. If, however, observers are asked to detect targets in a sequence of rapidly presented items, recognition of a target succeeding another target by about a third of a second is severely impaired, the "attentional blink" (AB) [Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology. Human Perception and Performance, 18*, 849–860]. Since most experiments on the AB use well controlled but artificial stimuli, the question arises whether the same phenomenon occurs for complex, natural stimuli, and if so, whether its specifics depend on stimulus category. Here we presented rapid sequences of complex stimuli (photographs of objects, scenes and faces) and asked observers to detect and remember items of a specific category (either faces, watches, or both). We found a consistent AB for both target categories but the duration of the AB depended on the target category.


*Keywords:* Attentional blink; RSVP; Natural scenes; Faces; Categorization

## 1. Introduction

When processing complex natural stimuli, humans grasp the "gist" of a scene within a small fraction of a second. This remarkable capability has often been probed using rapid serial visual presentation (RSVP) tasks. In an early demonstration, Potter and Levy (1969) presented series of images at rates between 0.5 and 8 Hz. After each of these RSVP sequences, subjects were asked to look through a set of images and to decide for each image whether it had been presented in the sequence. While the ability of subjects to recollect the scenes dropped with presentation speed, they still performed above chance at the highest tested rate (8 Hz). Biederman (1981) demonstrated that subtle violations of natural relations—such as a fire-hydrant standing on top of a mail box—are detectable in scenes, presented as briefly as 150 ms and followed by a mask. Coarse categorization of objects (e.g., animal vs. non-animal) in natural scenes is possible for stimuli displayed for only 20 ms (unmasked), though in these experiments the earliest category-dependent signal in the event-related potential (ERP) began about 150 ms after stimulus onset (Thorpe, Fize, & Marlot, 1996). All these findings highlight the remarkable processing speed of the human visual system, especially for complex natural stimuli.

When observers are instructed to respond to or remember a particular item ("target") in an RSVP sequence, the detection of a second target (T2) is impaired if it is presented in close succession (about 200–600 ms) after the first target (T1). This impairment, the so-called "attentional blink" (AB), is absent if T2 appears directly after T1 (Raymond, Shapiro, & Arnell, 1992). In their original report of the AB, Raymond et al. (1992) defined T1 by its color (a white letter in a sequence of black letters) and T2 by the

* Corresponding author. Present address: Institute of Computational Science, ETH Zurich, Switzerland. Fax: +41 44 632 1562.
*E-mail address:* wolfgang.einhaeuser@inf.ethz.ch (W. Einhäuser).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.12.007

occurrence of a particular exemplar (a black X). When the target is categorically defined (e.g., a letter among non-letters), the AB exhibits the same characteristic: no impairment for the item immediately following T1, but strong impairment for subsequent items (Chun & Potter, 1995). Based on these experiments, Chun and Potter (1995) put forward a two-stage RSVP model: in the first stage, items presented in a RSVP sequence are rapidly recognized and (coarsely) categorized, but are subject to fast forgetting unless they are consolidated in a further processing stage. If a target is detected in the first stage, a second, slower, and limited-capacity stage is initiated. When T2 directly follows T1, both targets enter the second stage. But if T2 falls within the period of the AB, it is processed in the first stage, but no second-stage processing is initiated since this stage is still occupied with processing T1. Hence T2 is rapidly forgotten. The two-stage concept of the AB has recently found support in event-related potential (ERP; Kranczioch, Debener, & Engel, 2003; Sergent, Baillet, & Dehaene, 2005) and functional magnetic resonance imaging (fMRI; Marois, Yi, & Chun, 2004) studies.

A critical feature of the two-stage model is the assumption of a common attentional "bottleneck" in the second-stage target processing. To have good control of the stimulus parameters, most studies of the AB used simple stimuli, such as single letters or symbols. However, it is unclear whether results obtained on such (seemingly) simple stimuli can be transferred directly to more natural conditions. Using a dual-task paradigm, Li, VanRullen, Koch, and Perona (2002) find that observers can classify natural stimuli into coarse categories (animal and vehicle) in the (near) absence of attention, whereas the classification of arbitrarily rotated letter stimuli fails under the same conditions.

Rousselet, Fabre-Thorpe, and Thorpe (2002) compare event-related potentials (ERPs) when two images are presented concurrently to a situation in which only one image is presented, while subjects perform the animal vs. non-animal go/no-go task. Consistent with their earlier study (Thorpe et al., 1996), they find that target and distractor ERPs start to diverge about 130 ms (occipital) or 160 ms (frontal) after stimulus onset. Differences between one- and two-image conditions, however, do not occur before 190 ms after stimulus onset. In addition, Rousselet et al. (2002) confirm Li et al.'s (2002) finding that behavioral performance is only slightly impaired in the two-image condition. Rousselet et al. demonstrate that this impairment is consistent with a simple model of parallel processing. Taken together the behavioral and ERP results indicate that early visual processing is highly parallelized and a presumed attentional bottleneck must occur late during processing. Besides supporting the notion of a late attentional bottleneck, these findings also raise the question on whether other attentional phenomena—such as the AB—differ between simple and natural stimuli.

Several studies have investigated the AB using natural stimuli for targets only, while employing scrambled versions of the same images as distractors or masks (Awh et al., 2004; Marois et al., 2004). In such a setting, Awh et al. (2004) found that T1 faces induced an AB for T2 letters, but not vice versa. Awh et al. argue that any account of the AB that assumes a *single and central* bottleneck is inconsistent with their results. Following their argument, this is irrespective of whether the bottleneck limits formation of working memory traces (as in the model of, e.g., Chun & Potter, 1995), limits availability of multiple items to "awareness for the control of behavior" (Duncan, Ward, & Shapiro, 1994) or limits the transition from visual short-term memory (VSTM) to retrieval (as in Shapiro, Caldwell, & Sorensen, 1997). Alternatively, Awh et al. (2004) suggest that there are multiple parallel stage-two resources. Only when the processing of T1 occupies all these resources an AB occurs for T2. Awh et al. (2004), however, use isolated stimuli, followed by a mask, at two different spatial locations. Whether their results transfer to a RSVP sequence of natural stimuli presented in a single location has remained unaddressed.

Recently, Evans and Treisman (2005) presented a series of natural scenes for 110 ms each to probe for an AB (their Experiments 4–7). In their case, animals and/or vehicles formed the target categories. When both types of targets had to be "identified", i.e., had to be classified into a subordinate category, AB increased "in depth and duration" when T1 and T2 belonged to different categories as compared to when T1 and T2 were within the same category. If T1 had to be only "detected", however, the AB shortened considerably. When both targets were in the same category but only had to be "detected," the AB was absent; when they were in different categories, it was strongly reduced. These results extend the "two-stage" model insofar as they constrain the demands for both stages. In particular, they are consistent with detection being largely supported by the first stage, whereas thorough identification requires the second stage. While this study differs from previous studies in using natural stimuli and distinguishing identification from detection, it leaves several AB issues open. First, Evans and Treisman (2005) presented at least one distractor between T1 and T2. Thus, they did not test the absence of the AB at short inter-target intervals. Second, they—as for most previous AB studies—used only one RSVP rate. This did not allow them to detect a difference in AB duration smaller than their chosen stimulus onset asynchrony (SOA). Finally, they defined identification as correct naming of the subcategory (vehicle type or animal species), but not as identification of a particular exemplar.

Here we presented subjects with 5-s RSVP sequences of natural stimuli at several rates between 6 and 40 Hz. To measure the full time-course of the AB, we placed 2 or 4 targets at random in the RSVP sequence, including short intervals between T1 and T2. The primary purpose of the four-target trials was ensuring subjects' persistent alertness throughout the sequence, even if two targets occurred early. We asked observers to remember all exemplars of the target category (faces, watches, or both—depending on