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

session). After the 5-s sequence, we tested target identification and memory two-alternative-forced-choice against a similar exemplar from the same category. We used the results to study how the time-course of the AB for identification/memorization depended on presentation rate and target category.

2. Methods

2.1. Subjects

Six volunteers from the Caltech community (age 20–32, two females) participated in the experiment. All subjects gave written informed consent to participation in the study and received payment for participation. The experiment conformed to national and institutional guidelines for experiments with human subjects and to the Declaration of Helsinki.

2.2. Setup

The experiment was conducted in a dark room specifically designed for psychophysical experiments. Stimuli were presented using a Matlab (Mathworks, Natick, MA) psychophysics toolbox extension (Brainard, 1997; Pelli, 1997) on a 19-inch CRT monitor (Dell Inc., Round Rock, TX). The monitor was set to a resolution of 1024×768 pixels and a refresh rate of 120 Hz. The subject viewed the stimuli from a distance of 100 cm. The 256×256 pixel-wide stimuli thus spanned about $6° \times 6°$ of visual angle. Maximum luminance of the screen ("white") was set to 25 cd/m^2; the ambient light level was below 0.04 cd/m^2.

2.3. Stimuli

As detailed below, the present study used two-target categories, faces and watches, embedded in a large variety of background stimuli. Face stimuli were taken from the "AR face database" (Martinez & Benavente, 1998; http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html) with permission of the authors. This database consists of frontal views of 131 different individuals (59 females), each of which photographed in 13 different configurations (different illumination, wearing sun-glasses or scarves) in two separate sessions. For the purpose of the present study we only used the 13 different configurations of the first session for each ID. Face stimuli were converted to 8-bit grayscale using Matlab's rgb2gray() function, resized to half the original resolution (384×288) using Matlab's imresize() function and cropped centrally to span 256×256 pixels.

Watch and background stimuli were obtained from the "Caltech-101" database (http://www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html; Fei-Fei, Fergus, & Perona, 2004), which contains a varying number of stimuli in 101 distinct object categories. In order to avoid accidental inclusion of the target category in the background stimuli, we excluded from the background set the categories "watches", "faces", "google background", "dollar-bill," and "Buddha," as well as images from other categories that contained human faces or watches. In addition, we excluded stimuli whose aspect ratio was larger than 1.5. Color images were converted to 8-bit grayscale, and were cropped to the length of the shorter dimension. The resulting square images were subsequently rescaled to 256×256 pixels. In total 1703 (=131*13) distinct faces, 239 distinct watches and 6890 distinct background stimuli were used in the experiments.

2.4. Protocol

Each subject participated in three experimental sessions. One of the sessions ("face session") used only faces as targets, another ("watch session") used only watches as targets, and the third ("dual-target session") used both watches and faces as targets. Each subject performed one session per experimental day. The order of sessions was balanced across the six subjects.

Each session consisted of 300 trials, half of which contained two targets, the other half-four targets. Targets were embedded at random temporal locations in 5-s RSVP sequences consisting mainly of randomly selected background stimuli (Fig. 1). While we here choose both target locations independently, which yields an abundance of short SOAs[1], for the investigation of AB balancing over SOAs presents an alternative. As we, however, are also interested in how overall identification and memorization performance depends on rate and category, and our analysis (see below) nevertheless allows investigation of the AB, there is no principle advantage of this alternative. Hence we decided to use independent choice of target locations (Obviously, we draw the temporal location without replacement, such that there are always two or four targets in the respective trials).

Each background stimulus occurred only once in an individual trial. The first and the last 500 ms of each 6 s presentation showed a fixation cross on a medium luminance (gray) background that spanned the same screen region as the stimuli. Stimuli were presented in direct succession (i.e., without blanks or masks between them) at six different rates (6, 12, 15, 20, 30, and 40 Hz), corresponding to SOAs from 25 to 167 ms per image. This gave 25 (300/(2*6)) trials per rate and condition. The order of trials within a session was random. Before each trial subjects were reminded, by a text slide, whether they had to "remember faces", "remember watches" or "remember faces and watches" in the current session.

Subjects began each trial by pressing a button. After the RSVP sequence, subjects were given two two-alternative-forced-choice (2-AFC) questions (Fig. 1), and indicated which face and/or watch (that shown on the left or right of the question image) they had seen in the preceding RSVP sequence. Subjects were instructed to respond "as accurately as possible and as fast as possible, without sacrificing accuracy."

To discourage the use of low-level cues, in the case of faces the distractor for the 2-AFC questions was chosen to have the same configuration (i.e., lighting, sun-glasses, scarf, etc.) as the target face. None of the face stimuli was used more than once per session. Since four-target trials primarily served to ensure constant alertness, even if two targets occurred early in the sequence, also in the four-target trials, only two targets were tested by 2-AFC questions. In four-target trials in dual-target sessions, one question per category, i.e., one face and one watch question, were posed. The order of questions ("Which face?" and "Which watch?") was random and did not relate to the order of their presentation in the RSVP sequence. Between the response to the first question and the onset of the second question there was an interval of 500 ms, in which the response of the subject was displayed. Auditory feedback was provided to the subjects as to the correctness of their decision for each 2-AFC question in each trial.

2.5. Data-analysis

2.5.1. Dense sampling of time-points

To investigate the attentional blink (AB), we assessed how recognition performance for the nth target presented in a trial, T(n), depended on the time interval between the onset of T(n) and that of the onset the preceding target T($n-1$). This interval hereafter will be referred to as the SOA between T(n) and T($n-1$). To obtain a dense sampling of AB latencies, we analyzed target SOAs across all presentation rates used. We first sorted all data for a given subject and session by SOA, then averaged the proportion of correct responses for targets within different overlapping SOA windows.

[1] Target positions within each RSVP sequence were drawn randomly from a uniform distribution, i.e., a target could occur in each frame with equal probability. The absolute difference between two random variables drawn with uniform probability from the same finite set, peaks at 0 and decreases monotonically towards larger differences. Hence there were more short target SOAs than large target SOAs, allowing a particularly dense sampling of low target SOAs, of particular interest for AB research.

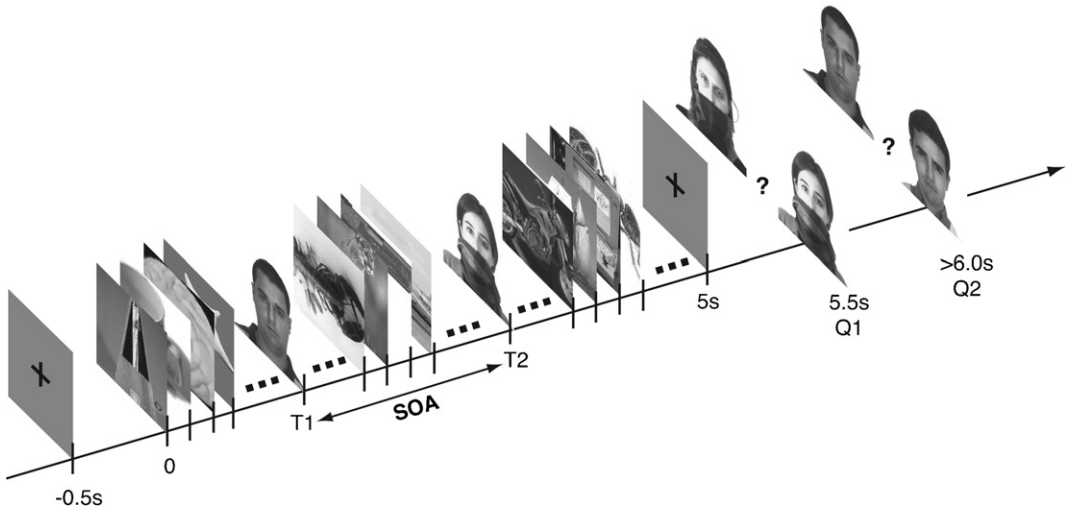

Fig. 1. Task design. Rapid presentation sequence begins 500 ms after subject presses any key ($t = 0$). Two or four targets (faces, watches, or faces and watches) are embedded in a rapid serial visual presentation (RSVP) sequence at random latencies. Subjects are tested on the identification of two targets using two 2-AFC questions. In the case of faces, the distractor in the 2-AFC shared the same configuration (sun-glasses, scarf, lighting) as the target.

Since all targets were placed randomly in each trial, small SOAs were more frequent than large ones. To obtain a roughly constant number of samples in each SOA-bin, we increased the width of the performance smoothing-window logarithmically. Each window spanned one latency octave (i.e., the lower latency limit was half the upper limit) and successive windows were advanced in 0.0027-octave (the minimum occurring difference between two SOAs) steps. Performance within each such SOA window was counted as representing performance at the window's arithmetic-mean log SOA or equivalently, at its geometric-mean SOA. This analysis is performed first individually for each subject. All figures that depict performance over SOA show the mean and standard error across subjects of these SOA-smoothed performances.

2.5.2. Statistical analysis/surrogate datasets

To have sufficient amount of data for robust analysis in each SOA-bin, all statistical analysis is performed on the SOA-smoothed data. To test whether we have a significant AB in a given SOA-bin, we perform a *t*-test with the null hypothesis that performance in this bin is identical to the mean performance for SOAs larger than 1.5 s[2] in the same condition. To test whether two conditions (watch vs. face) significantly differ from each other in a given time bin, we perform a *t*-test with the null hypothesis that performance is identical in both conditions. In all cases we use the raw SOA-smoothed performance of each of the six individual subjects as the independent samples for the *t*-tests, without any prior normalization.[3]

Since the targets were placed randomly in each RSVP sequence, small target SOAs were dominated by higher presentation rates. Any effect that shows differences between small and large SOA windows therefore requires verification that it is not attributable to this bias. To control for this potential confound and for the fact that we perform a large number of individual *t*-tests without correction for multiple comparisons, we generated surrogate data: For each subject and presentation rate, we randomly reassigned the target SOAs of one trial to the performance (correct/incorrect) of another trial. This procedure keeps the aggregate performance and distribution of SOAs *for each rate* constant. Any result that is attributable just to the biases in SOAs and overall performance would also occur in the surrogate analysis. Any effect found in the original, but not in the surrogate data, therefore must be a consequence of the relation between SOA and performance. We repeated this random remapping 100 times and performed exactly the same analyses on each of the 100 surrogate data sets as on the original data. Values of interest were averaged over these 100 surrogate sets at each time-point. The presence of a significant difference between different SOAs in the surrogate data would be indicative of a statistical (sampling) artifact. If, however, the original data did, but the surrogate data did not show any significant effect, the observed effect in the original data cannot be attributed to a sampling artifact. Hence the surrogate data served a control for the validity of the performed analysis.

3. Results

3.1. Overall performance

First we analyzed the dependence of recognition performance on presentation rate, independent of the time of occurrence of the target in the RSVP sequence. For face targets, we found, as expected, a strong anti-correlation between presentation rate and recognition performance ($r = -.94$, $p = .005$, Fig. 2a). The same held as well for watches ($r = -.92$, $p = .01$, Fig. 2b) and also in "dual-target" sessions in which both faces and watches were targets ($r = -.89$, $p = .02$, Fig. 2c). Despite the decrease in performance with increasing RSVP rate, mean recognition performance was above chance (50%) for all tested rates and target types (Fig. 2a–c). A *t*-test for individual rates revealed that the recognition performance was significantly (at $p < .05$) above chance for all tested rates and categories, with the exception of watch targets presented at 40 Hz

[2] Obviously the choice of 1.5 s as boundary for "large" SOA is somewhat arbitrary. We chose 1.5 s as about half the trials (913/1800) have SOAs beyond this boundary and most earlier studies implicitly assume the AB to be over at such large SOAs. The SOA for which exactly half of the trials have larger SOA, half of the trials have smaller SOA is 1.53 s.

[3] There would be several possible schemes by which one could normalize within subject before analysis. Normalizing with respect to average performance would potentially over-represent small SOAs, normalizing with respect to the mean of SOA-smoothed representation would over-represent long SOAs. Finally, one could normalize to performance at large SOAs, but as the choice of what presents a "large" SOA is somewhat arbitrary, we decided to stick with the raw performances for analysis.

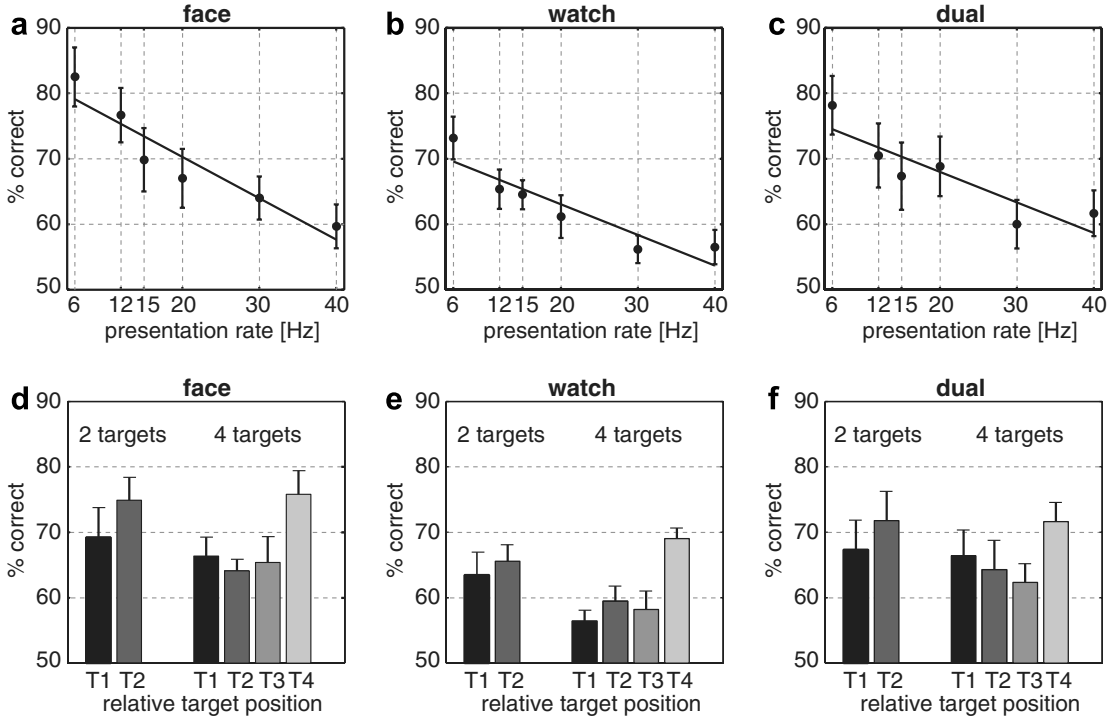

Fig. 2. Overall performance. (a–c) Performance as a function of presentation rate for (a) face-target sessions (b) watch-target sessions and (c) dual-target sessions. Mean and standard error of the mean (SEM) across subjects are displayed. The line corresponds to best linear fit. See Table 1 for significance relative to chance level. (d–f) Performance as a function of the relative position of target in the RSVP sequence for (d) face-target sessions (e) watch-target sessions and (f) dual-target sessions. Error bars denote SEM across subjects.

(Table 1). These results show that—while recognition performance degraded with increasing presentation rate—observers recognized targets even at rates as high as 30 or 40 Hz, i.e., for image presentations as short as 25 ms.

Next we analyzed whether recognition performance depended on the number of targets per trial, on the latency of the target in the RSVP sequence, or on whether the target was the first (T1), second (T2), third (T3), or fourth (T4) in the RSVP sequence. For all target categories, both target numbers (2 and 4), and for all latencies of the targets within the RSVP sequence, recognition performance was significantly (at $p = .05$, t-test) above chance (Fig. 2d–f, Table 2). There is a trend for T2 targets to be better recognized than T1 targets in two-target trials (Fig. 2d–f). However, this trend was not significant for any target category (faces: $p = .39$; watches: $p = .66$; dual: $p = .53$). A similar trend appeared in four-target trials, in which T4 was better recognized than T1–T3 for all target categories. Pairwise comparisons between T4 and T1, T2, or T3 were significant only for watches (Fig. 2e, T1–T4: $p = .006$; T2–T4: $p = .01$; T3– T4: $p = .01$), while for faces (Fig. 2d, T1–T4: $p = .09$; T2–T4: $p = .02$; T3–T4: $p = .11$) and dual-target sessions (Fig. 2f, T1–T4: $p = .35$; T2–T4: $p = .23$; T3–T4: $p = .07$) no consistent effects were observed. It seems likely that these trends were not a consequence of better recognition of the last target in each trial, but rather reflect a slight recency effect by which targets appearing later in the sequence were remembered better during the ensuing question period.

Table 1
Performance for different presentation rates

	6 Hz	12 Hz	15 Hz	20 Hz	30 Hz	40 Hz
Face	0.0008	0.0014	0.0094	0.0129	0.0080	0.0345
Watch	0.0009	0.0038	0.0013	0.0187	0.0334	0.0556
Dual	0.0015	0.0087	0.0200	0.0090	0.0429	0.0207

p-values for t-tests of null hypothesis that performance equals chance.

Table 2
Performance for different relative positions of targets in the RSVP sequence (T1–T4)

	Two-target trials		Four-target trials			
	T1	T2	T1	T2	T3	T4
Face	0.0117	0.0013	0.0036	0.0008	0.0162	0.0013
Watch	0.0170	0.0025	0.0169	0.0133	0.0460	0.0001
Dual	0.0171	0.0068	0.0131	0.0328	0.0106	0.0011

p-values for t-tests of null hypothesis that performance equals chance.

In conclusion, despite the expected strong dependence of performance on presentation rate, plus a possible slight dependence on the latency or relative position of the targets in the RSVP sequence, observers achieved above-chance performance for all target latencies, target categories, and nearly all presentation rates used.

3.2. Attentional blink within category

As a first analysis of the attentional blink (AB), we analyzed trials in which exactly two targets of the same

category (either faces or watches) were presented. In Fig. 3a we plot the percentage of correct responses to the second target (T2) as a function of its SOA to the preceding target (T1). For faces, the grand mean across subjects of this moving-window performance index exhibits a sharp dip to close to chance levels for SOAs near 300 ms—the AB.

To assess the significance of the AB, we compared performance in each latency window to mean performance for large target SOAs (SOA > 1.5 s). By this definition, across subjects the AB was significant ($p < .05$) for latency windows centered between 236 and 377 ms as well as for a short period between 401 and 406 ms (Fig. 3b). Identical analysis on the surrogate data (see Section 2) showed no significance at any time-point (the minimum across all time-points was $p_{\min,sur} = .34$; see Fig. 3b, dotted line). This ensures that the observed AB was not a statistical artifact and hence confirms a significant AB for face targets, presented in a 140-ms latency window centered at about 300 ms.

Performance on the target category of watches displayed qualitatively similar behavior (Fig. 3c). The AB, however, was significant (again at $p < .05$, in comparison to mean performance level for SOAs > 1.5 s) over a larger and later range of target SOAs (356–371, 377–424, 442–471 and 566–589 ms; Fig. 3d). Again the surrogate data shows no region of significant difference ($p_{\min,sur} = .41$; Fig. 3d, dotted line). While these data confirm an AB for watch targets, they suggest that the AB occurred later and lasted longer for watch targets than for face targets.

To test whether the AB duration indeed depended on stimulus category, we directly compared performance for watch and face targets for each target SOA latency window. This direct comparison showed performance to be

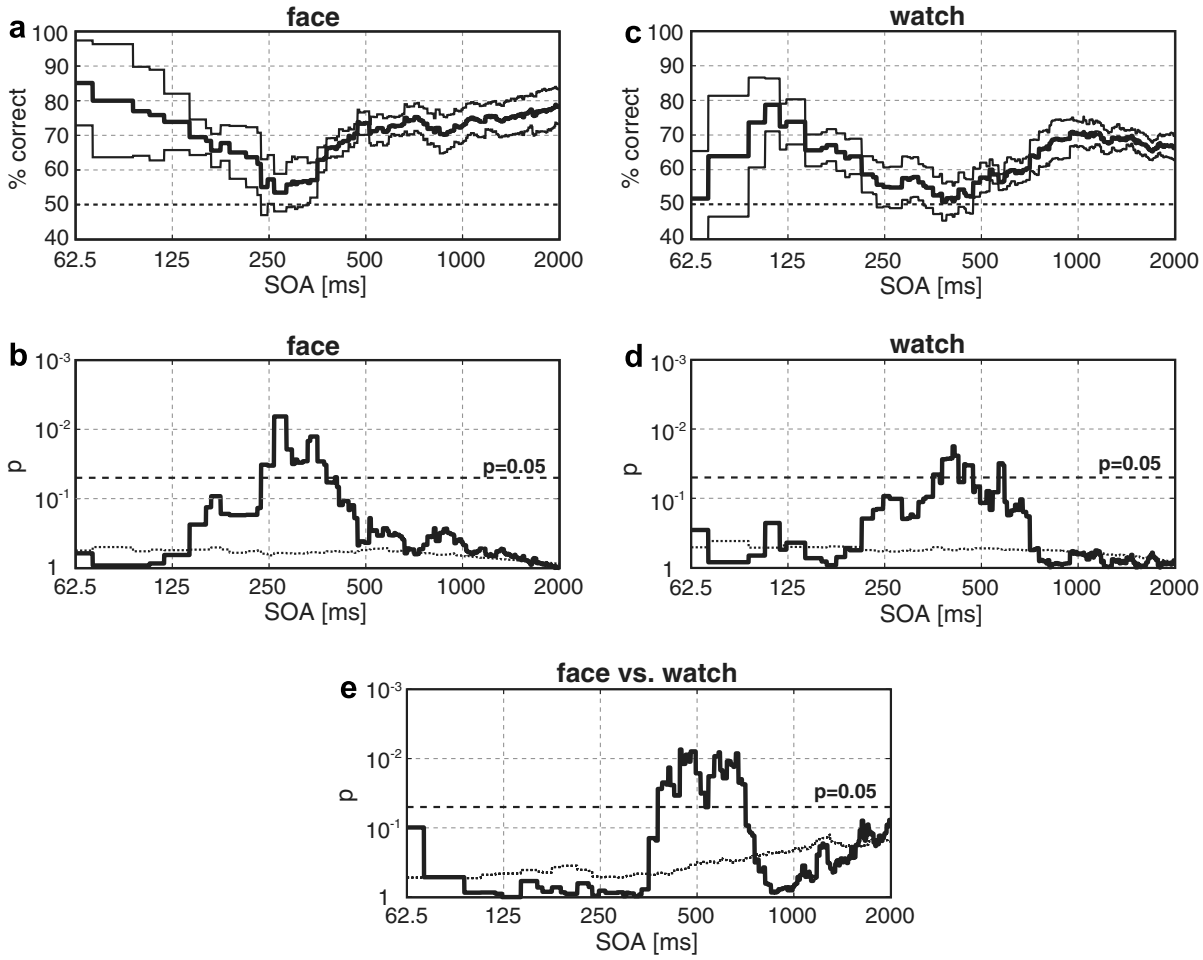

Fig. 3. Attentional blink in single-category sessions, two-target trials. (a) Performance recognition of the second target in a trial (T2) as a function of the SOA between T1 and T2 in face-target sessions. Thick line shows the mean across subjects, thin lines denote the mean ± standard error at each SOA. Note that the time axis is scaled logarithmically. (b) Statistical difference (*t*-test) between face-target recognition performance in each SOA window compared to mean performance at long inter-target SOAs (>1.5 s). Both axes are scaled logarithmically, with higher significance levels towards the top. The dashed line indicates the $p = .05$ level; the dotted line denotes the $p = .05$ result of surrogate analysis performed to control for statistical artifacts. (c) Performance for T2 as a function of SOA between T1 and T2 for watch-target only sessions. Notation as in (a). (d) Statistical difference between watch-target recognition performance in each SOA window compared to mean performance at long inter-target SOAs (>1.5 s). Notation as in (b). (e) Statistical significance of difference between watch-target and face-target T2-recognition performance as a function of SOA. Notation as in (b).

significantly worse for watch targets than for face targets for SOAs between 377 and 530 ms as well as between 542 and 707 ms ($p < .05$ by t-test at each SOA latency; Fig. 3e), but not at any longer target SOAs. Within these significance ranges, mean performance was 15% (percentage points) worse for watch targets than for face targets. Again the observed significance could not be explained by a statistical artifact, as the surrogate data did not show a significant effect at any SOA range ($p_{\min,\mathrm{sur}} = .12$; Fig. 3e, dotted line). The range of SOAs in which performance differs significantly between target categories indicates a category dependence of the AB.

3.2.1. Four-target trials

The original rationale to embed four-target trials within in the two-target trials was to ensure constant alertness throughout the trial, even when two targets occurred early in the RSVP sequence. Hence we queried recognition for only two targets in the four-target trials. Direct comparison to the two-target situation would be confounded by the reduced amount of data. Nevertheless, we could analyze the AB in trials with four targets, by plotting recognition performance for target Tn (for $n = 2, 3, 4$) in dependence of the SOA between Tn and T($n - 1$). For face targets we do not observe a significantly (at $p = .05$ in comparison to performance at SOA > 1.5 s) AB-like Tn performance drop at short SOAs between Tn and T($n - 1$) (Fig. 4a and b). Part of this lack of effect might be attributable to a putative AB between T($n - 2$) and T($n - 1$) interfering with the perception of T($n - 1$) and therefore diminishing its effect on Tn. For watch targets, however, we observe a significant AB between Tn and T($n - 1$) (Fig. 4c). The SOA ranges in which AB was significant for watches included 248–377 and 413–448 ms (Fig. 4d). This result further indicates a category dependence of the AB. Direct comparison confirmed a difference in AB between the

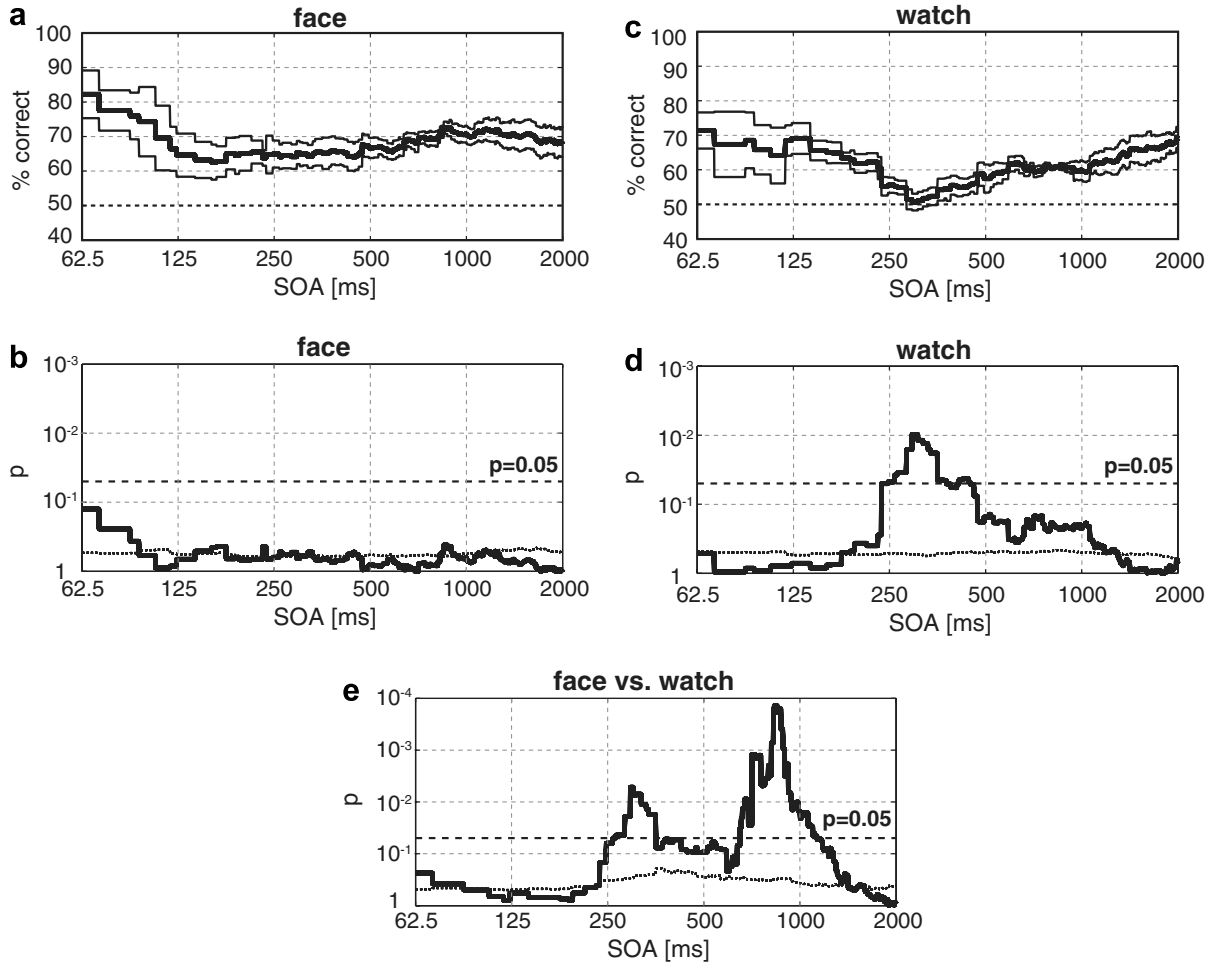

Fig. 4. Attentional blink in single-category session four-target trials. (a) Performance on the nth target ($n = 2,3,4$) in four-target trials plotted as a function of SOA between nth and ($n - 1$)th target in face-targets only sessions. Notations as in Fig. 3a. (b) Statistical difference (by t-test) between face-target Tn recognition performance in each SOA window compared to mean performance for long SOAs (>1.5 s). Notations as in Fig. 3b. (c) Performance on nth target ($n = 2, 3, 4$) in four-target trials plotted as a function of SOA between nth and ($n - 1$)th target in watch-targets only sessions. Notation as in Fig. 3a. (d) Statistical difference (by t-test) between watch-target recognition performance in each SOA window compared to mean performance at long SOAs (>1.5 s). Notation as in Fig. 3b. (e) Statistical significance of difference between watch-target and face-target recognition performance in each SOA window. Notation as in (b).

two categories: performance was significantly worse for watches than for faces between 265 and 353 ms and between 649 ms and 1.13 s (Fig. 4e). Mean performance difference in these ranges was 11% (percentage points).

At long SOAs, however, not even a trend towards a performance difference was observed (minimum probability, for SOAs larger than 1.5 s, was $p = .68$, Fig. 4e). This renders it unlikely that the target category difference in the AB resulted from a general performance difference in recognizing faces as compared to watches. Again the surrogate analysis did not show any significance in any SOA window ($p_{\min,\mathrm{sur}} = .20$, Fig. 4e, dotted line). In conclusion—while there are several differences between four-target and two-target trials, and while performance for Tn in four-target trials may be influenced by other targets than T($n-1$) alone, the main observation that AB depends on stimulus category, also prevails for four-target trials.

3.3. Attentional blink across categories

The observed difference in AB duration between categories might arise from the nature of the stimulus inducing the AB (T1) or from the nature of the (un)recognized stimulus itself (T2). Using the "dual-target sessions" allowed us to distinguish these alternatives. If the T2 category plays the dominant role, the recognition performance for T2 should not depend on the category of T1. With faces as T2 and watches as T1, we found a less pronounced and later occurring AB than for faces as both T2 and T1 (Fig. 5a). Statistical analysis revealed that the performance for T2 faces in the dual-target session was significantly better than in the face-only session for SOAs in a range of intervals (236–247, 259–318, and 330–353 ms; Fig. 5b). For watches as T2, the same trend was observed: when T1 were faces, T2 watches were better recognized at shorter SOAs than if both targets were watches (Fig. 5c). This difference was significant for SOAs in ranges 319–389, 460–471, 519–530, and 566–589 ms (Fig. 5d). Surrogate analysis again verified that the observed significant difference was not a statistical artifact (face $p_{\min,\mathrm{sur}} = .49$, watch, $p_{\min,\mathrm{sur}} = .44$). At short SOAs for both target categories, T2 performance was better if T1 was of the other category than if T1 was of the same category. Consequently the AB is not only dependent on the category of the recognized target (T2), but also depends on the relation between the category of T2 and T1, with improved performance if categories are different.

An analogous analysis for the four-target trials (between Tn and T($n-1$)) did not reveal significant differences between watches as Tn and faces as T($n-1$) compared to the condition in which both Tn and T($n-1$) were watches ($p_{\min} = .06$). For faces as Tn, a difference was only observed only at very short SOAs (<48 ms) and for large SOAs (>1.8 s); this is unlikely to be related to the same AB phenomenon described above and for short SOAs might in part be attributed to task-switching effects between categories. Despite the absence of a consistent effect in the four-target trials, the two-target trials provide strong evidence that the observed difference between stimulus categories depends not only on the category of the

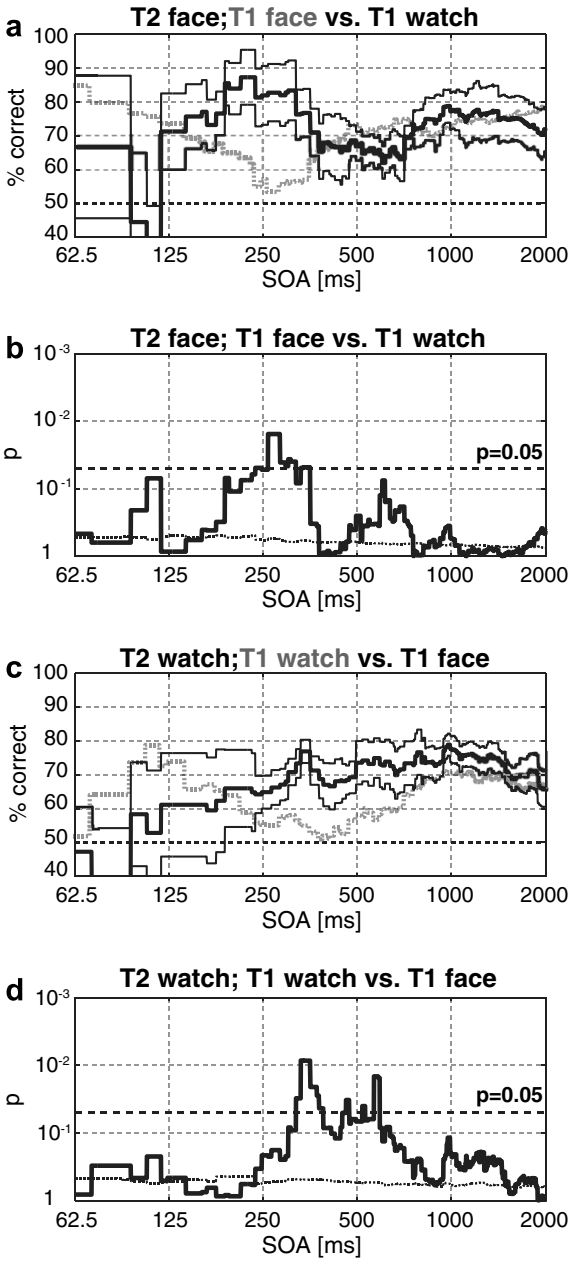

Fig. 5. Comparison between single and dual-target category trials. (a) Black lines denote performance on T2 face targets in dual-target sessions as a function of SOA between T2 (face) and T1 (watch), thick line denotes mean, thin lines (mean ± SEM) across subjects. Dotted gray line plots result of Fig. 3a (T1 and T2 face) for comparison. Axes as in Fig. 3a. (b) Significance of difference in face recognition performance between face-targets only and dual-target sessions. Notation as in Fig. 3b. (c) Black lines denote performance on watch targets in dual-target sessions as a function of SOA between T2 (watch) and T1 (face); thick line show the mean, thin lines the mean ± SEM across subjects. Dotted gray line plots the result shown in Fig. 3c (T1 and T2, watch targets) for comparison. Axes as in Fig. 3a. (d) Significance of difference in watch-target recognition performance between watch-target only and dual-target sessions. Notation as in Fig. 3b.

target, whose recognition the AB impairs (T2), but also on whether or not T1 and T2 are of the same category.

4. Discussion

Here we demonstrate an attentional blink (AB) for the recognition and memorization of natural stimuli. We show that the duration of the AB depends on target category.

The AB has been described for a variety of stimuli and conditions. The first report (Raymond et al., 1992) defined T1 by color and T2 by a particular letter in a letter sequence. Using digits, letters and symbols, Chun and Potter (1995) showed that categorically defined targets also cause an AB. Attentional deficits for processing subsequent targets have also been reported for stimuli such as colors (Ross & Jolicoeur, 1999) and words (Luck, Vogel, & Shapiro, 1996). Furthermore, even "preattentive" processing in a dual task (Joseph, Chun, & Nakayama, 1997) was impaired at short lags between targets. There is also evidence that the processing of auditory stimuli can impair subsequent visual attention, indicative of a cross-modal AB (Jolicoeur, 1999).

While all these studies show an impairment of attentional or at least of target processing, many of these studies did not report the second hallmark of the original AB description: near-normal level of performance for very small inter-target SOAs ("lag-1 sparing"). Since this feature typically distinguishes the AB from other types of attentional impairments, such as repetition blindness (Kanwisher, 1987; for dissociation from AB, see Chun, 1997) and task-switching costs (see Potter, Chun, Banks, & Muckenhoupt, 1998), the investigation of short lags is crucial for testing the AB.

Here we observed, using natural photographic stimuli, an AB time-course that strongly resembles the AB as originally described by Raymond et al. (1992) in their letter task: performance showed little reduction for targets closely following the first target and a dip in performance for inter-target SOAs near 300 ms. The duration of the AB found in the aforementioned studies varies considerably; both of our conditions—face targets and watch targets — are well within the previously described ranges. However, as most previous studies only used one or very few presentation rates, the observed AB durations are neither directly comparably across studies, nor were those studies designed to reveal small differences in AB duration. By using a variety of presentation rates ranging from 6 to 40 Hz, we achieve a fine-grained mapping of the time-course of the AB that allowed us to uncover differences in the AB time-course between different target categories.

Several models of attention have been suggested to account for the AB (Chun & Potter, 1995; Duncan et al., 1994; Giesbrecht & Di Lollo, 1998; Raymond et al., 1992; Raymond, Shapiro, & Arnell, 1995; Shapiro, Raymond, & Arnell, 1994; Shapiro et al., 1997). While they differ in the detailed locus of the capacity limitation, and debate early vs. late stages in processing as well as limitations of working memory consolidation vs. limitations in maintaining object representations, all these models are built upon the assumption of a central attentional resource, whose limitation in capacity causes the AB. However, in the case of natural stimuli, there are indications that at least coarse categorical processing takes place in the (near) absence of attention (Li et al., 2002; Rousselet et al., 2002). Indeed, in one of the few studies on the AB that used naturalistic stimuli, Awh et al. (2004) found evidence against a *single* central attentional resource. In particular, face identification (on a small set of potential targets) was not impaired by a digit discrimination task, while the same task induced a strong AB for letters. Based on these findings, Awh et al. suggest a "multi-channel" (i.e., parallel) account of the AB, which requires all channels to be occupied by processing T1 before an attentional blink occurs. Such a multi-channel model offers one possible interpretation for the category dependence of the AB profile we observe here, if one assumes that faces and watches use (partly) different channels. Nevertheless, several mechanisms are conceivable that may influence the temporal profile of T2 recognition, even after the T1-induced bottleneck has ended. Whereas further investigation of such potential mechanisms is beyond the scope of the present paper, the herein observed category dependence constrains future models of the AB.

Unlike in studies using single stimuli that are masked after presentation, in RSVP two different effects might impair performance at high presentation rates: not only is the presentation duration of each individual stimulus reduced, but there are also more stimuli to be processed in the same amount of time. To dissect these influences and measure short SOAs at constant presentation durations, Potter, Staub, and O'Connor (2002) presented words at different spatial locations. For a presentation duration of 53 ms, they found that at short SOAs (17–53 ms), T2 was more likely to be correctly reported than T1, at 107 ms T1 and T2 showed about equal performance, and only at higher SOAs (213 ms in their study) T2 performance was impaired relative to T1 performance, i.e., the typical AB was observed. Based on this result, Potter etal. suggest an account of lag-1 sparing based on a two-stage model for the AB: T1 and T2 compete for resources of a first processing stage; if the recognition task is sufficiently difficult and T2 follows very shortly after T1 (very short SOA) the competition leads to impaired processing of T1. For increasing SOAs, however, T1 more and more benefits from its "head-start" in the competition, which shifts the advantage towards T1, impairing T2 thus causing the AB. Only the stimulus winning successfully competing for resources in the first stage can then be processed in the second stage, i.e., be consolidated in visual short-term memory. While in the present study presentation duration also varies with SOA, our results are consistent with two important predictions of the Potter et al. (2002) model. First, rather than only sparing T2 at "lag-1", the performance gradually decreases with SOA and reaches its minimum only at about SOAs of 300 ms. Second, performance is better if T1 and

T2 are of different categories as compared to the same category. In the framework of Potter et al.'s (2002) results and model, our data indicate that at the first stage processing resources for faces and watches are only partly overlapping. Consequently the AB can show different time-courses for faces and watches, even if the second stage presents a unique and central bottleneck. The Potter et al. (2002) model thus can reconcile the herein observed category dependence of the AB in natural scenes with the notion of a central, late and unique bottleneck.

In a recent study, Evans and Treisman (2005) tested for the AB using natural stimuli. Their observers had to *detect* targets defined by category (vehicle, animal). In some of their experiments, observers in addition had to *identify* (*and memorize*) the target by naming its subordinate category. These authors found a difference between detection and identification: while target detection alone showed nearly no AB, the AB for identification was comparable with the effect described here. Since our task also required target identification, the fact that we find an AB is in general agreement with their findings. In addition, Evans and Treisman found that the AB for target identification lasted longer than for simple stimuli; their results were well in the range of the AB observed in our watch target task.

In their Experiment 4, Evans and Treisman observed a difference in overall identification performance: observers identified animals better than vehicles. However, they did not observe a significant interaction between target SOA, task condition, and target category, and concluded that, "the same attentional effects appear with both target types", in contrast to our results. There are several potential causes for this apparent discrepancy: (i) We used different categories, and it is possible that the difference in attentional effects between animals and vehicles is considerably smaller than that between faces and watches. Of course, other psychophysical differences between the sets of images used might also explain the difference. (ii) Evans and Treisman used an indirect statistical measure—interaction via an ANOVA—to support their conclusion, while we directly measured the difference between categories that were otherwise (i.e., for longer inter-target SOAs) equally well identified. (iii) They used only a single—comparably low—presentation rate (8 Hz), which might leave a latency difference of the extent reported here unnoticed. (iv) They had subjects identify the target subcategory freely, while we imposed a forced choice decision. Whether the discrepancy in the results arose from differences in task, stimuli or presentation rate is an interesting issue for further research. More important than the differences in experimental details, however, both studies agree that there are differences in AB duration between simple and natural stimuli. We furthermore show that there are also natural stimulus categories between which AB duration may differ.

In conclusion, we here report that the duration of the AB depends on target category. This result constrains models of the AB, which need to account for this category dependence. Our findings also underline the care that has to be taken when generalizing results obtained using simple stimuli to predictions about visual performance under more natural conditions.

Acknowledgments

This work was financially supported by the Swiss National Science Foundation (WE: PBEZ2—107367; PA00A-111447), the National Institutes of Health USA, the National Science Foundation, and by DARPA/NGA.

References

Awh, E., Serences, J., Laurey, P., Dhaliwal, H., van der Jagt, T., & Dassonville, P. (2004). Evidence against a central bottleneck during the attentional blink: multiple channels for configural and featural processing. *Cognitive Psychology, 48*(1), 95–126.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kuboby & J. R. Pomerantz (Eds.), *Perceptual organisation*. Hillsdale: NJ Lawrence Erlbaum Associates.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance, 21*(1), 109–127.

Chun, M. M. (1997). Types and tokens in visual processing: a double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology. Human Perception and Performance, 23*(3), 738–755.

Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature, 369*(6478), 313–315.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology. Human Perception and Performance, 31*(6), 1476–1492.

Fei-Fei, L., Fergus, R., & Perona, P. (2004). Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision.

Giesbrecht, B., & Di Lollo, V. (1998). Beyond the attentional blink: visual masking by object substitution. *Journal of Experimental Psychology. Human Perception and Performance, 24*(5), 1454–1466.

Jolicoeur, P. (1999). Restricted attentional capacity between sensory modalities. *Psychonomic Bulletin & Review, 6*(1), 87–92.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a 'preattentive' feature search task. *Nature, 387*(6635), 805–807.

Kanwisher, N. G. (1987). Repetition blindness: type recognition without token individuation. *Cognition, 27*(2), 117–143.

Kranczioch, C., Debener, S., & Engel, A. K. (2003). Event-related potential correlates of the attentional blink phenomenon. *Brain Research. Cognitive Brain Research, 17*(1), 177–187.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Science USA, 99*(14), 9596–9601.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*(6601), 616–618.

Martinez, A. M., & Benavente, R. (1998). The AR Face Database. CVC Technical Report.

Marois, R., Yi, D. J., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron, 41*(3), 465–472.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10–15.

Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: the visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology. Learning, Memory and Cognition, 24*(4), 979–992.

Potter, M. C., Staub, A., & O'Connor, D. H. (2002). The time course of competition for attention: attention is initially labile. *Journal of Experimental Psychology. Human Perception and Performance, 28*(5), 1149–1162.

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.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1995). Similarity determines the attentional blink. *Journal of Experimental Psychology. Human Perception and Performance, 21*(3), 653–662.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5*(7), 629–630.

Ross, N. E., & Jolicoeur, P. (1999). Attentional blink for color. *Journal of Experimental Psychology. Human Perception and Performance, 25*(6), 1483–1494.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience, 8*(10), 1391–1400.

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1994). Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. *Journal of Experimental Psychology. Human Perception and Performance, 20*(2), 357–371.

Shapiro, K. L., Caldwell, J., & Sorensen, R. E. (1997). Personal names and the attentional blink: a visual "cocktail party" effect. *Journal of Experimental Psychology. Human Perception and Performance, 23*(2), 504–514.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522.