Decomposing the Attentional Blink

Petersen, Anders; Vangkilde, Signe

Published in:
Journal of Experimental Psychology: Human Perception and Performance

DOI:
10.1037/xhp0001018

Publication date:
2022

Document version
Peer reviewed version

Citation for published version (APA):
Decomposing the Attentional Blink

Anders Petersen¹ & Signe Vangkilde¹

¹ Department of Psychology, University of Copenhagen

Word Count: 7739

Author Note

Anders Petersen  [ORCID: https://orcid.org/0000-0003-3595-6275]
Signe Vangkilde [ORCID: https://orcid.org/0000-0002-1622-3108]

This research was supported by a grant from the Independent Research Fund Denmark; Are you in control? The cognitive control and arousal project (CoCoA), grant number 9037-00169B.

Correspondence concerning this article should be addressed to Anders Petersen, Institute of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 Copenhagen K. Email: anders.petersen@psy.ku.dk. Data and study materials are available upon request.
Abstract

The Attentional Blink (AB) refers to a deficit in reporting a second target (T2) embedded in a stream of distractors when presented 200-500 ms after a preceding target (T1). Several theories about the origin of the AB have been proposed; filter-based theories claim that the AB is the result of a temporarily closing of an attentional gate to avoid featural confusion for targets and distractors, while bottleneck theories propose that the AB is caused by a reduction in the capacity to either encode into or maintain information in visual short-term memory. In three experiments, we systematically vary the exposure duration and composition of the T2 display allowing us to decompose the T2 deficit into well-established parameter estimates based on the Theory of Visual Attention (TVA). As the different AB theories make specific predictions regarding which parameters should be affected during the AB, we are able to test their plausibility. All three experiments consistently show a lower capacity to process T2 during the AB, supporting theories hypothesizing a bottleneck at the encoding stage. No evidence is found supporting filter-based theories or theories placing the bottleneck at the maintenance stage.

Keywords: Attentional blink, Temporal visual attention, Theory of visual attention, Computational modelling, Visual processing capacity.

Public Significance Statement: This study uses computational modelling of the Attentional Blink (AB) task to quantify the temporal limitations of visual attention in humans. Results consistently show that it is a temporary reduction in visual processing capacity which gives rise to the AB.
Decomposing the Attentional Blink

Visual attention is limited and allows only a fraction of the information in our visual field to reach conscious awareness. To investigate the temporal dynamics of this limitation, researchers have used a paradigm in which two targets are embedded in a rapid serial visual presentation (RSVP) stream of distractors. Results show that participants can correctly report the first target, but not the second target, if the two targets are separated by 200-500 ms. This phenomenon is known as the attentional blink (AB). A number of theories and computational models have been proposed to explain the AB (see Dux & Marois, 2009, and Martens & Wyble, 2010, for a thorough reviews). The theories generally fall into two main groups: filter-based theories and bottleneck theories.

Filter-based theories

This first group of theories suggests that the AB occurs when a mechanism designed to filter out irrelevant information is temporarily disturbed and suppresses any incoming stimuli irrespective of their relevance. Raymond et al. (1992) proposed the first filter-based theory, the Inhibition model. The theory suggests that the AB is caused by the closing of an attentional gate to reduce target and distractor featural confusion. Thus, when the distractor immediately following T1 is presented, the gate closes and is not opened until T1 is fully identified resulting in the reduced ability to correctly identify T2 when presented in close temporal proximity of T1. Olivers and Meeter (2008) revised and extended the Inhibition model in their computational Boost and Bounce model, which assumes that the gating system is used to enhance relevant and suppress irrelevant information. When an item in the stream is a target, a transient excitatory feedback mechanism will boost the sensory representation of the target and the following items in the stream. The boost is strongest for the item presented immediately after the target. Thus, if T2 directly follows T1, it will be spared and T2 performance is not impaired. However, if the item following T1 is a distractor, the boost will trigger a strong transient inhibitory feedback mechanism (the bounce) causing the input filter (the gate) to block following items in the stream from being encoded into visual short-term memory (VSTM). Similar to a boost, a bounce will be strongest for the next item in the stream. Thus, an AB will occur if T2 is presented at Lag 2 (i.e., two positions after T1), but will be gradually attenuated for longer lags. A similar emphasis on the role of top-down target selection processes is found in the temporary loss of control theory.
(Di Lollo et al., 2005). In this theory, the AB is a result of a temporary loss of control over an endogenous filter that selects targets and excludes distractors. The presentation of T1 makes the filter vulnerable to stimulus-driven disruption from distractors. If T1 is followed by a distractor, the filter will no longer allow targets to enter and an AB is observed. The dynamics of attentional selection also plays a central role in the *delayed attentional engagement account* (Nieuwenstein et al., 2005). According to this account, the AB is a consequence of a delayed reengagement of attention to T2 shortly after it has been disengaged from T1 (to avoid processing of subsequent irrelevant information). As a whole, the filter-theories link the AB closely to selective attention and how attention is adjusted to avoid distractor interference.

**Bottleneck theories**

The second group of theories suggests that the AB arises because of a limited capacity mechanism (a bottleneck) along the processing pathway which makes visual information available for conscious awareness. Chun and Potter (1995) proposed a highly influential bottleneck theory, the *two-stage model*. In Stage 1, stimuli are recognized, but their representations are vulnerable to both decay and overwriting by subsequent stimuli. In Stage 2, stimuli are encoded/consolidated into VSTM. Stage 2 is capacity limited as only one stimulus can be processed at a time. According to the two-stage model, the AB is a result of T2 being held in the unstable Stage 1 until T1 has finished processing in Stage 2.

Inspired by Chun and Potter’s two-stage limited-capacity model, Bowman and Wyble (2007; see also Wyble et al., 2009) proposed a formal computational model, the *Simultaneous Type, Serial Token* (STST) model. In Stage 1, a parallel visual processing of the stimuli is performed up to the level of semantic categorization (type representation). For a stimulus to enter VSTM it must be bound to a token in VSTM that provides episodic information about where the stimulus was located in the stream. This binding happens in Stage 2 by activation of a transient attentional enhancement (TAE), facilitating the bottom-up representation of a stimulus. According to the STST model, the AB occurs because the TAE is temporally suppressed after activation until T1 has been bound to a token and consolidated into VSTM.

An alternative bottleneck theory was proposed by Ward et al. (1996). In their *attentional dwell-time hypothesis*, the AB is a result of a competition between T1 and T2 for limited processing resources. The head start of T1 increases its probability of winning the competition, leaving T2 open to interference from
trailing stimuli. The very first bottleneck theory, the *interference theory* by Shapiro et al. (1994) did not, however, place the bottleneck in the encoding stage, but rather in VSTM. The interference theory assumes that stimuli entering VSTM will be assigned a weighting based on the available space in VSTM and their similarity to the selection (target) templates defined by the current task. According to the interference theory the AB occurs because both targets (T1 and T2) and their distracting successors (T1+1 and T2+1) all enter VSTM, resulting in T2 only receiving a diminished weighting and, thus, being more open to interference from the other items in VSTM.

**Theory of Visual Attention (TVA)**

In this article, we test some of the assumptions on which the above AB theories rely by decomposing the T2 deficit into well-established parameter estimates based on the Theory of Visual Attention (TVA; Bundesen, 1990). TVA provides a formal computational framework for making quantitative analyses of attentional functions. TVA describes the visual encoding and selection process as a capacity-limited parallel processing race in which all stimuli compete simultaneously for representation in a VSTM store with limited capacity. Processing capacity is distributed according to attentional weights, so that objects of high importance are processed faster and thereby have a higher probability of capturing one of the few slots in VSTM. According to TVA, the course of the visual encoding and selection process depends on the following distinct parameters: $K$, the capacity of VSTM measured in number of letters; $C$, the visual processing capacity measured in letters processed per second; $t_0$, the threshold of conscious perception or the longest ineffective exposure duration measured in seconds; $\alpha$, the top-down controlled selectivity defined as the ratio between the attentional weight of a distractor and the attentional weight of a target. $\alpha$ values close to 0 indicate efficient selection of targets and values close to 1 indicate no prioritization of targets compared with distractors.

TVA-based modelling of T2 performance requires a systematic variation of T2 exposure duration. The majority of AB studies have used a constant presentation rate of around 10 stimuli/s, resulting in an effective exposure duration of 100 ms for T2 (e.g., Chun & Potter, 1995). Several studies have investigated the AB for shorter and longer exposure durations of T2. Giesbrecht & Di Lollo (1998; see also McLaughlin et al., 2001; Bowman & Wyble, 2007) showed that the AB persists for shorter exposure durations as long as
T2 is masked by an interruption mask, whereas for longer exposure durations the AB was diminished. Jannati et al. (2012), however, convincingly argued that ceiling effects may explain the absence of the AB for longer exposure durations and reported evidence that the AB is in fact invariant with respect to T2 exposure duration. This was further supported by a study by Jolicœur and Dell’Acque (2000) in which the effects of AB magnitude and T2 exposure duration are reported to be additive. In sum, this is evidence that variation of T2 exposure duration can be used as a tool to investigate the AB without interfering with the AB per se.

In Experiment 1, we use TVA-based modelling to investigate how different gating (primarily used by filter-based theories) vs. bottleneck mechanisms may give rise to the T2 deficit. The experiment uses a standard AB paradigm with the addition that the exposure duration of T2 is systematically varied before T2 is terminated by a pattern mask. Figure 1 outlines three scenarios of how different gating vs. bottleneck mechanisms can lead to a T2 deficit when presented at Lag 3 (i.e., during the AB) compared with Lag 8 (i.e., outside the AB). Scenario I reflects an all-or-none gating mechanism which with a certain probability, $p_{gat}$, closes a gate before T2 onset and keeps it closed for the entire duration of T2 presentation. This will result in a proportionally lower probability of correctly reporting T2 at Lag 3 across all exposure durations. Scenario II reflects a delayed gating mechanism which closes a gate before T2 onset, but reopens it before T2 offset. This scenario will manifest itself as a longer ineffective exposure duration, $t_0$, at Lag 3 compared with Lag 8. Finally, Scenario III reflects a bottleneck mechanism allowing T2 to be processed for its entire duration but with a reduced processing capacity, $C$.

In Experiment 2, T2 displays are composed of six targets (whole-report displays) and participants are to report both T1 and as many targets from the T2 display as possible. Based on these data we are able to estimate the same set of parameters as in Experiment 1, but also the VSTM capacity, $K$, of the T2 display, making it possible to investigate whether the AB is a result of a capacity limitation in the encoding stage and/or the maintenance stage.

In Experiment 3, we include both T2 displays with only targets (whole-report displays) and T2 displays containing targets and distractors (partial-report displays). By combining these display types, it is possible to estimate the same set of parameters as in Experiments 1 and 2, as well as the selectivity
parameter, $\alpha$ (see Vangkilde et al., 2011 for details about the CombiTVA setup). Comparing the selectivity during and outside the AB is of interest since filter-based theories claim a temporal loss of control of the input filter which basically closes off all input from the RSVP stream. If so, this should result in similar attentional weights of targets and distractors leading to an increase in the selectivity parameter, $\alpha$, where higher $\alpha$-values indicate poorer selectivity.

In all three experiments, we include one or two control conditions to investigate the impact of the initial distractor stream on the TVA-parameters. Experiment 1 includes a condition in which a single target with varying exposure duration is presented after the initial distractor stream (Stream condition), whereas Experiments 2 and 3 include both a condition with and without (No-stream condition) the initial distractor stream presented before the multi-item display. In Experiment 1, we expect no difference in target processing when targets are presented before (Stream condition) and after the AB (Lag-8 condition). In Experiments 2 and 3, we expect the presence of the initial distractor stream to improve selectivity (lower $\alpha$-values). Evidence for such a pre-target selectivity state in the AB paradigm has previously been reported by Dux et al. (2006) and Maki and Padmanabhan (1994).
Figure 1

Possible scenarios for the function of different gating vs. bottleneck mechanisms

Note. Three possible scenarios for the function of different gating vs. bottleneck mechanisms: I) An all-or-none gating mechanism which with a certain probability, $p_{\text{gate}}$, closes a gate before T2 onset and keeps it closed for the entire duration of T2 presentation. II) A delayed gating mechanism which closes a gate before T2 onset, but reopens it before T2 offset. III) A bottleneck mechanism which reduces the processing capacity of T2. The red and green lines represent T2 performance as a function of exposure duration, when T2 is presented at Lag 3 (during the AB) or at Lag 8 (outside the AB), respectively. For each of the three scenarios, the white, red, and green boxes represent the hypothetical hazard rates (height of boxes; the conditional probability density that the target is encoded into VSTM at a given timepoint provided that the target has not been encoded before that timepoint) as a function of exposure duration (length of boxes) for T1, T2 presented at Lag 3, and T2 presented at Lag 8, respectively. The black boxes represent the closing of a gate. The scenarios predict: I) a higher $p_{\text{gate}}$ (i.e., $1-p_{\text{gate}}$ is the asymptote of the curve), II) a longer $t_0$ (i.e., the point at which the curve rises from the abscissa), and III) a lower $C$ (i.e., the slope of the curve at $t_0$), respectively, during the AB.
Experiment 1

In Experiment 1, participants reported the identity of either two target letters (T1 and T2) or a single target letter embedded in a RSVP stream of distracting digits. T1 was presented for the same exposure duration as the digits, whereas T2 was presented for varying exposure durations, making it possible to perform a TVA-based modeling of T2 performance. T2 was presented either at Lag 3 or Lag 8 following T1, or as T1 in a single target control condition. TVA-based modeling of the data was performed for each individual subject and independently for the three conditions.

Method

Sample size selection

The sample size required to observe a significant effect for the Lag-3 vs. Lag-8 comparison was determined based on a study by Grassi et al. (2020), a replication of the original AB study by Raymond et al. (1992) with N = 98 participants. A robust effect was found with an effect size of $d_z = 1.38$. Using G*Power 3.1 (Faul et al., 2013) with an alpha of .05 and power of .95, a minimum sample size of nine was determined. Because of the desire to analyze data using mathematical modeling, we opted for a substantially larger sample size of 24 participants in Experiment 1 and 26 participants in Experiments 2 and 3.

Participants

24 Danish students (19 females, 5 males, mean age = 25 years) were paid a standard fee per hour for participating in the experiment. All had normal or corrected-to-normal vision. The study was approved by the institutional ethical review board (No. IP-IRB / 17062020), but not preregistered. Data and study materials are available upon request.

Design

In two thirds of the trials, T2 was presented equally often at either Lag 3 or Lag 8 following T1, whereas in the remaining one third of the trials T2 was presented as T1 (Stream condition; see Figure 2). The temporal location of the first target was uniformly distributed between the 7th and 14th location in RSVP stream. Target letters were chosen randomly without replacement from a set of 20 letters (ABCDEFGHJKLMPRSTUXYZ), such that the identity of T1 and T2 differed and such that each letter was
presented once and only once as T1 and T2 for each exposure duration and condition. Exposure durations were chosen from a set of six durations (20, 50, 80, 140, 200, and 300 ms). The factorial design resulted in one test session of 360 trials (3 conditions × 6 exposure durations × 20 letters).

Procedure

Stimuli were presented on a 21" CRT monitor running at 100 Hz using E-prime 2 software with participants seated approximately 60 cm from the monitor in a semidarkened room. The beginning of a trial was signaled by a black fixation cross (0.6° × 0.6° of visual angle, 0.1 cd/m²) presented centrally on a grey background (52.8 cd/m²) for a duration of 1000 ms followed directly by the RSVP stream presented on the same central location. The stream comprised a series of distracting digits and one or two embedded target letters. All stimuli were black (0.1 cd/m²) and presented in the Electra font for which alphanumeric characters are made up of small black squares placed on a grid covering five squares in the horizontal direction (0.9° of visual angle) and seven squares in the vertical direction (1.2° of visual angle). Each digit in the stream was presented for 100 ms and randomly chosen from a set of 9 digits (1 to 9) with the restriction that consecutive digits had to differ. T1 was always presented for 100 ms and replaced by the next digit in the stream, whereas T2 had a variable exposure duration and was masked by a black pattern mask. Masks were constructed by randomly placing black squares in half of the locations in a grid of seven (horizontal, 1.2° of visual angle) by nine (vertical, 1.5° of visual angle) locations. In total, 19 different masks were constructed and used at random with replacement. If a mask was constructed with a very uneven distribution of black squares, it was discarded and a new mask was constructed to ensure that all masks would be equally effective. The selected mask was shown for 500 ms. Hereafter, participants made a nonspeeded report of the identity of the letter. Participants were instructed to only report letters they were “fairly certain” of having seen (i.e., to use all available information but refrain from pure guessing). To keep participants motivated, a 500 ms feedback display informed participants of whether their report of each presented letter was correct (‘*’) or not (‘-’). Participants completed 40 practice trials and 1080 test trials (i.e, 3 × 360 trials) divided into blocks of 40 trials. Participants were encouraged to take small breaks between blocks.
Figure 2

Trial outline for Experiment 1

![Trial outline for Experiment 1](image)

Results

Figure 3 shows the mean proportion of correctly reported T2 for each exposure duration of T2 and for each of the three conditions. Only trials in which T1 was correctly reported were included in the analysis of the Lag-3 and Lag-8 conditions. For each individual participant and condition, TVA-parameters were estimated by a maximum-likelihood procedure using the LibTVA toolbox (Dyrholm et al., 2011). The model had three parameters: perceptual threshold ($t_0$; 1 degree of freedom [df]), visual processing capacity ($C$; 1 df), and probability of a closed gate ($p_{gate}$; 1 df)$^1$. Table 1 shows the mean for each parameter estimate with a margin of error at 95% confidence. Model fits to data from a representative participant is presented in Figure S1 (Supplementary material). Two-tailed paired-sample $t$ tests with a bonferroni corrected alpha-level of $0.05/3 = 0.017$ revealed a significant drop in processing capacity, $C$, for T2 presented at Lag 3 compared with

---

$^1$ For single-item displays (Experiment 1), the probability that a letter was correctly reported at exposure duration $t$ was modelled as: $p(t) = 0$ for $t < t_0$ and $p(t) = (1-p_{gate})(1-exp[-C(t-t_0)])$ for $t \geq t_0$. Note that $p_{gate}$ was estimated by specifying the mathematically equivalent model in LibTVA that included the probability of making an attentional lapse (originally intended for research into ADHD, McAvinue et al., 2015). For details on how to model multi-item displays (Experiments 2 and 3), we refer the interested reader to Dyrholm et al. (2011; see also Shibuya & Bundesen, 1988).
Lag 8, $t(23) = 4.28, p < .001, d_z = 0.87$. No significant difference was found for either $t_0$ or $p_{gate}$ when comparing the Lag-3 and Lag-8 conditions, nor did any of the TVA-parameters differ significantly when comparing the Stream and Lag-8 conditions. Model selection based on the Bayesian Information Criterion (BIC) showed that Scenario III (summed BIC = 9113) should be preferred over Scenario I (summed BIC = 9202) and II (summed BIC = 9479). Additionally, a slightly lower T1 performance was observed at Lag 8 ($M = .77$) compared with Lag 3 ($M = .79$), $t(23) = 3.78, p = .001, d_z = 0.77$. In sum, the results of Experiment 1 are in favor of Scenario III (see Figure 1) in which a bottleneck mechanism allows T2 to be processed for its entire duration but with a reduced processing capacity, $C$.

**Figure 3**

*Results for Experiment 1*

![Graph showing the mean proportion of correctly reported letters for different exposure durations and conditions.](image)

*Note.* The figure shows the mean proportion of correctly reported targets (with 95% confidence intervals) for each exposure duration, when only a single target was presented in the stream of distractors (black), when T2 was presented at Lag 3 (red), and when T2 was presented at Lag 8 (green). Only trials in which T1 was correctly reported were included for the Lag-3 and Lag-8 conditions.
Table 1

Means of parameter estimates with margins of error at 95% confidence and effect sizes (Cohen’s $d_z$)

<table>
<thead>
<tr>
<th></th>
<th>T2</th>
<th>T1 Lag 3</th>
<th>T2</th>
<th>T1 Lag 8</th>
<th>$d_z$</th>
<th>Stream</th>
<th>No-stream</th>
<th>$d_z$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1 ($n = 24$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_0$</td>
<td>41.44 ± 10.30</td>
<td>33.32 ± 6.13</td>
<td>0.29</td>
<td>27.13 ± 5.95</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td><strong>12.04 ± 2.20</strong></td>
<td>21.06 ± 3.14**</td>
<td>0.87</td>
<td>21.30 ± 2.94</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{gate}$</td>
<td>0.015 ± 0.013</td>
<td>0.006 ± 0.006</td>
<td>0.28</td>
<td>0.007 ± 0.006</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 2 ($n = 23$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_0$</td>
<td>21.63 ± 10.10</td>
<td>13.13 ± 4.46</td>
<td>0.40</td>
<td>8.85 ± 2.68</td>
<td>18.23 ± 7.09*</td>
<td>0.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td><strong>15.75 ± 4.38</strong></td>
<td>21.68 ± 3.51**</td>
<td>0.78</td>
<td>21.90 ± 3.21</td>
<td>26.72 ± 4.05</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{gate}$</td>
<td>0.033 ± 0.031</td>
<td>0.004 ± 0.004</td>
<td>0.43</td>
<td>0.002 ± 0.003</td>
<td>0.006 ± 0.006</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K$</td>
<td>2.49 ± 0.15</td>
<td>2.51 ± 0.16</td>
<td>0.07</td>
<td>3.13 ± 0.30</td>
<td>3.04 ± 0.26</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experiment 3 ($n = 25$)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_0$</td>
<td>14.25 ± 6.38</td>
<td>12.70 ± 4.93</td>
<td>0.09</td>
<td>7.90 ± 2.44</td>
<td>13.61 ± 5.49*</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td><strong>17.96 ± 2.67</strong></td>
<td>28.69 ± 3.46**</td>
<td>1.98</td>
<td>28.73 ± 3.71</td>
<td>31.13 ± 3.92</td>
<td>0.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{gate}$</td>
<td>0.005 ± 0.008</td>
<td>0.001 ± 0.001</td>
<td>0.20</td>
<td>0.000 ± 0.000</td>
<td>0.000 ± 0.000</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K$</td>
<td>2.66 ± 0.14</td>
<td>2.59 ± 0.12</td>
<td>0.42</td>
<td>3.17 ± 0.20</td>
<td>3.05 ± 0.21</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.80 ± 0.17</td>
<td>0.61 ± 0.09</td>
<td>0.51</td>
<td>0.62 ± 0.06</td>
<td>0.75 ± 0.07**</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p*-value < alpha-level (Exp. 1 and 2: alpha-level = .05/3 = .017; Exp. 3: alpha-level = .05/4 = .0125; Exp. 5: alpha-level = .05/5 = .01), **p*-value ≤ .001. Significant differences for Lag-3 and Lag-8 conditions are presented in bold.
Experiment 2

In Experiment 2, we tested the hypothesis that the lower performance at Lag 3 is a result of a VSTM limitation. By extending the T2 displays to include six targets (whole-report displays), it is possible to estimate the same set of parameters as in Experiment 1, but also the VSTM capacity, $K$, of the T2 display. More specifically, $K$ is the asymptotic level of the number of correctly reported letters if the exposure duration of the T2 display is increased to infinity (cf. Figure 5). As a further extension of Experiment 1, we included a condition in which the whole-report displays were presented without the initial distractor stream (i.e., No-stream condition). Comparing this to the condition in which the whole-report displays were preceded by a distractor stream without T1 (i.e., Stream condition), we are able to examine the direct influence of the initial distractor stream on target performance.

Method

Participants

26 Danish students (15 females, 11 males, mean age = 25 years) were paid a standard fee by the hour (6 students) or received course credits (20 students) for participating in the experiment. All had normal or corrected-to-normal vision.

Design

The design was similar to that of Experiment 1 with the addition of a condition in which the whole-report displays were presented without the initial distractor stream (No-stream condition; see Figure 4). In all conditions, the whole-report displays were presented for seven different exposure durations (10, 20, 50, 80, 140, 200, and 300 ms). Target letters were chosen randomly without replacement from the same set of 20 letters as in Experiment 1, such that on every trial the identity of T1 and the letters in the whole-report display differed. Each test block consisted of $2 \times 28$ trials ($4 \text{ conditions} \times 7 \text{ exposure durations}$).

Procedure

The procedure was similar to that of Experiment 1 except that T2 consisted of a whole-report display with six targets presented equidistantly on an imaginary circle with a radius of 2.3° of visual angle. As in Experiment 1, participants made a nonspeeded report of the identity of the letters. Participants were
instructed to report the letters they were “fairly certain” of having seen and aim for a response accuracy between 80% and 90% (i.e., the percentage of correctly reported letters out of all reported letters). After each block, participants were informed about the accuracy of their responses. Participants completed two practice blocks of 16 trials [4 conditions × 4 exposure durations (50, 140, 200, and 300 ms)] and 15 test blocks of 56 trials.

Figure 4

Trial outline for Experiment 2

Results

Three participants only reported T1 correctly in less than 15% of the trials in the Lag-3 and Lag-8 conditions and were thus excluded. Figure 5 shows the mean of the participants’ mean number of correctly reported letters for each exposure duration of the whole-report displays and for each of the four conditions. Only trials in which T1 was correctly reported were included in the analysis of the Lag-3 and Lag-8 conditions. Individual TVA-parameters were estimated for each participant and condition. The model had four parameters: perceptual threshold ($t_0$; 1 df), visual processing capacity ($C$; 1 df), probability of a closed...
gate \(p_{gate}; 1 \text{ df})\), and short-term memory capacity \((K; 5 \text{ df}; \text{the } K \text{ value reported is the expected } K \text{ given a particular distribution of the probability that on a given trial } K = 1, 2, \ldots, 6)\). Table 1 shows the mean for each parameter estimate with a margin of error at 95% confidence. Model fits to data from a representative participant is presented in Figure S2 (Supplementary material). Two-tailed paired-sample \(t\) tests with a bonferroni corrected alpha-level of \(0.05/4 = 0.0125\) revealed a significant drop in processing capacity, \(C\), for T2 presented at Lag 3 compared with Lag 8, \(t(22) = 3.75, p = 0.001, d_z = 0.78\). No significant differences were found for \(t_0, p_{gate}\) or \(K\) when comparing the Lag-3 and Lag-8 conditions. Further, model selection based on the Bayesian Information Criterion (BIC) showed that a model with equal \(K\)-values for the Lag-3 and Lag-8 conditions (summed BIC = 25726) should be preferred over a model with separate \(K\)-values (summed BIC = 26329). A likelihood-ratio test for nested models was in agreement with this result, \(\chi^2(115) = 54.3, p > 0.9999\). Comparisons of the parameter estimates for the No-stream and the Stream conditions showed a significantly lower \(t_0\) in the latter condition, \(t(22) = 2.81, p = 0.01, d_z = 0.59\). No significant differences were found for \(C, p_{gate}\) or \(K\) when comparing these conditions. The mean proportion of correctly reported T1 in the Lag-3 condition \((M = .63)\) and the Lag-8 condition \((M = .64)\) did not differ significantly, \(t(22) = 2.00, p = 0.058, d_z = 0.42\). In sum, the results replicate the findings in Experiment 1: Only processing capacity, \(C\), was found to be significantly reduced during the AB. No reduction in VSTM capacity, \(K\), of the T2 display was observed.
Figure 5

Results for Experiment 2

Note. The figure shows the mean of the participants’ mean number of correctly reported letters (with 95% confidence intervals) for each exposure duration of the whole-report displays, when displays were presented with (black) and without (blue) the initial distractor stream, and when displays were presented at Lag 3 (red) and Lag 8 (green). Only trials in which T1 was correctly reported were included for Lag-3 and Lag-8 conditions.
Experiment 3

In Experiment 3, we extended the design used in Experiment 2 to include a combination of T2 displays with six and two target letters (whole-report displays) and T2 displays with two target letters and four distractor digits (partial-report displays; see Vangkilde et al., 2011 for details about the CombiTVA paradigm). The difference in target performance between the two-target condition and the two-target-four-distractor condition reflects the participants’ ability to filter out the distractors and focus their attention on the targets (cf. Figure 7b). This ability can be quantified by estimating the selectivity parameter, $\alpha$, defined as the ratio of the attentional weight of a distractor on a certain spatial location in the display to the attentional weight of a target on the same location.

Method

Participants

26 Danish students (21 females, 5 males, mean age = 23 years) were paid a standard fee for participating in the experiment. All had normal or corrected-to-normal vision.

Design

The design was similar to that of Experiment 2 with the addition of multi-item displays containing two target letters alone or two target letters and four distractor digits (see Figure 6). The two targets were either presented both in the left hemifield, both in the right hemifield, or one in each hemifield. Distractors were chosen randomly without replacement from a set of nine digits (1 to 9) with the restriction that the four digits differed. All multi-item displays with two targets were presented for 100 ms, whereas multi-item displays with six targets were presented for seven different exposure durations as in Experiment 2 (10, 20, 50, 80, 140, 200, and 300 ms). Each test block consisted of 52 trials [4 conditions $\times$ (7 six-target display exposures + 3 two-target displays + 3 two-target-four-distractor displays)]$^2$.

---

$^2$ The design of the multi-item displays was equivalent to the design used in the CombiTVA paradigm (Vangkilde et al., 2011) except for the addition of a 300 ms exposure duration for the six-target displays and a slight increase in the exposure duration of the two-target displays from 80 ms to 100 ms. A pilot study showed that this modified design worked well with the Electra font.
Procedure

The procedure was similar to that of Experiment 2. Participants were instructed to make a nonspeeded report of the letters they were “fairly certain” of having seen and to ignore the digits. Participants completed two practice blocks of 24 trials [4 conditions × (3 six-target display exposures [50, 140, and 300 ms] + 1 two-target display + 2 two-target-four-distractor displays)] and 27 test blocks of 52 trials.

Figure 6

Trial outline for Experiment 3

Results

One participant only reported T1 correctly in less than 1.5% of the trials in the Lag-3 and Lag-8 conditions and was thus excluded. For each of the four conditions, Figure 7 shows the mean of the participants’ mean number of correctly reported letters for each exposure duration of the multi-item displays containing: a) six targets and b) two targets alone vs. two targets with four distractors. Only trials in which T1 was correctly reported were included in the analysis of the Lag-3 and Lag-8 conditions. For each
participant and condition, five TVA-parameters were estimated: perceptual threshold \((t_0; 1 \text{ df})\), visual processing capacity \((C; 1 \text{ df})\), probability of a closed gate \((p_{\text{gate}}; 1 \text{ df})\), VSTM capacity \((K; 5 \text{ df})\), and selectivity \((\alpha; 1\text{ df})\). Table 1 shows the mean for each parameter estimate with a margin of error at 95% confidence. Model fits to data from a representative participant is presented in Figure S3 (Supplementary material). Two-tailed paired-sample \(t\) tests with a bonferroni corrected alpha-level of \(0.05/5 = 0.01\) revealed a significant drop in processing capacity, \(C\), for T2 presented at Lag 3 compared with Lag 8, \(t(24) = 9.90, p < .001, d_z = 1.98\). No significant differences were found for \(t_0, p_{\text{gate}}, K,\) or \(\alpha\) when comparing the Lag-3 and Lag-8 conditions. Further, model selection based on the Bayesian Information Criterion (BIC) showed that a model with equal \(\alpha\)-values for the Lag-3 and Lag-8 conditions (summed BIC = 47314) should be preferred over a model with separate \(\alpha\)-values (summed BIC = 47444). A likelihood-ratio test for nested models was in agreement with this result, \(\chi^2(25) = 29.7, p = .238\). Comparisons of the parameter estimates for the No-stream and the Stream conditions showed a significantly lower \(t_0, t(24) = 3.08, p = .005, d_z = 0.62\), and an improved selectivity \((\alpha), t(24) = 3.76, p = .001, d_z = 0.75\), in the Stream condition. No significant differences were found for \(C, p_{\text{gate}}\) or \(K\) when comparing these conditions. The mean proportion of correctly reported T1 in the Lag-3 condition \((M = .72)\) and the Lag-8 condition \((M = .72)\) did not differ significantly, \(t(24) = 0.37, p = .712, d_z = 0.07\). In sum, the results replicate the findings in Experiment 1 and 2: Only the processing capacity, \(C\), was significantly reduced during the AB. The central stream of distractors affected T2 performance by lowering the perceptual threshold, \(t_0\), but also by enhancing selectivity (i.e., lower \(\alpha\)).
Figure 7

Results for Experiment 3

Note. The figure shows the mean of the participants’ mean number of correctly reported letters (with 95% confidence intervals) for each exposure duration of the multi-item displays with (a) six letters and (b) two targets alone, 2T, vs. two targets with four distractors, 2T4D, when displays were presented with (black) and without (blue) the initial distractor stream, and when displays were presented at Lag 3 (red) and Lag 8 (green). Only trials in which T1 was correctly reported were included for Lag-3 and Lag-8 conditions.

General Discussion

In three experiments, we used TVA-based modelling to decompose the T2 deficit during the AB into well-established attentional parameter estimates. In Experiment 1, we investigated which of three scenarios best explain the deficit (see Figure 1): In Scenario I, the deficit is a result of a gate being closed with a certain probability for the entire duration of T2 presentation; in Scenario II, the deficit is a result of a gate being closed at the onset of T2, but reopening before T2 offset; and in Scenario III, the deficit is a result of a bottleneck mechanism allowing T2 to be processed for its entire duration but with a reduced processing capacity. Experiment 1 used a standard AB paradigm with the addition that the exposure duration of T2 was systematically varied before T2 was terminated by a pattern mask. The results favored Scenario III as we
observed a significant reduction in the processing capacity of T2, C, but no significant increase in $p_{\text{gate}}$ (Scenario I) or $t_0$ (Scenario II). Model selection confirmed this result. Thus, we did not find evidence for an all-or-none gating mechanism (Scenario I), preventing T2 from being encoded into VSTM. This is particularly interesting since many current models of the AB rely on exactly such a mechanism but have to combine it with trial-by-trial variation in stimulus strength for the models to work (e.g., the boost and bound model, Olivers & Metter, 2008; and the STST model, Bowman & Wyble, 2007). Neither did we find evidence of a delayed gating mechanism (Scenario II) as predicted, for example, by the delayed attentional engagement account (Nieuwenstein et al., 2005). Based on our findings, we rather suggest that future models of the AB phenomenon incorporate a mechanism that lowers the likelihood of T2 being encoded into VSTM for all stimulus types and durations (e.g., by reducing the processing speed; Scenario III), and not a mechanism that sets the likelihood to zero for some stimulus types or durations as is custom in models relying on all-or-none and delayed gating mechanisms, respectively. In Experiment 1 we found that T1 performance was slightly lower in the Lag-8 condition ($M = .77$) compared with the Lag-3 condition ($M = .79$). One explanation for this finding could be that T1 needs to be retained longer in VSTM when T2 is presented at Lag 8 compared with Lag 3, resulting in a higher chance that the memory trace of T1 is lost before report. This is unlike most standard AB paradigms in which the length of the RSVP stream is the same across conditions (e.g., Chun & Potter, 1995). The finding was not replicated in any of the following experiments and should thus be treated with caution.

In Experiment 2, we replicated the main finding from Experiment 1 in a whole-report experiment with T2 consisting of six-target displays. That is, we found that only the processing capacity, C, was significantly reduced during the AB. Importantly, the experimental setup in Experiment 2 enabled the estimation of the VSTM capacity, K, of the T2 display during and outside the AB. According to the interference theory (Shapiro et al., 1994), the T2 display should receive a diminished weighting (less space) in VSTM when presented during the AB because both target displays (T1 and T2) and their distracting successors (T1+1 and T2+1) all enter and take up space in VSTM. The results of Experiment 2, however, showed no significant difference in VSTM capacity of the T2 display during and outside the AB, and model selection show that a model with equal K-values should be preferred. This finding speaks against the
hypothesis that the AB is a result of bottleneck and interference in the maintenance stage. A similar
collection was reached by Akyürek and Hommel (2005). They found that when memory load was high the
overall performance in the RSVP task was impaired, but the size of the AB remained unaffected by load.

Based on these findings, Akyürek and Hommel (2005) suggested that the AB reflects a capacity limitation in
the encoding of information into VSTM and not interference within VSTM.

In Experiment 3, we replicated the findings of Experiments 1 and 2. Again, we found that only the
processing capacity, $C$, was significantly reduced during the AB. Experiment 3 combined whole- and
partial-report displays and, thus, extended the set of estimated parameters to also include a selectivity
parameter, $\alpha$, (i.e., the efficiency of filtering out distractors and focusing attention on targets). Although
descriptively higher $\alpha$-values (i.e., poorer selectivity) were observed during the AB (Lag-3 condition)
compared with outside the AB (Lag-8 condition), this difference was not significant, and model selection
showed that a model with equal $\alpha$-values should be preferred. This poses a challenge for the filter-based
theories. According to these theories top-down target selection should be severely impaired during the AB
as attentional control over the input filter is temporarily lost (e.g., Di Lollo et al., 2005). In line with our
findings, several other studies have reported that attentional control is intact during the AB (e.g.,
Nieuwenstein, 2006; Zivony & Lamy, 2016). For instance, Nieuwenstein (2006) reported that successful
cuing during the AB requires the cue to match the target selection criteria (not just a feature of T2) and,
thus, that attentional control over selection criteria is maintained and intact during the AB. Similarly,
Zivony and Lamy (2016) showed that a stimulus matching the target selection criteria captures spatial
attention to the same extent whether it appears within or outside the AB, again indicating that attentional
control is intact during the AB. However, not all studies agree; Jiang and Chun (2001) showed that selection
of T2 during the AB becomes more difficult and is subject to larger spatial distractor interference,
suggesting that selection mechanisms are indeed compromised. And evidence from event-related potentials
(ERP; see Zivony and Lamy, 2021, for a recent review of ERP studies of the AB) studies shows that the
N2pc component (i.e., an ERP component linked to selective attention) is largely reduced during the AB
(Jolicoeur et al., 2006). Using a well-defined measure of selectivity (indexed by the $\alpha$-parameter)
Experiment 3 contributes to this discussion by showing only a small non-significant increase in distractor interference during the AB. Instead, the impaired performance is explained by a highly significant reduction in the processing capacity (indexed by the $C$-parameter) replicating the results from Experiments 1 and 2. In this context, it should be noted that TVA-parameters have a high reliability compared with scores derived from other attention tasks (Habekost et al., 2014), with the reliability of $\alpha$ being the lowest (Foerster et al., 2019). One possible explanation for the lower reliability of $\alpha$ is that the parameter is defined as a ratio between two variables (the attentional weight of a distractor and a target, respectively), a formal property that is known to make mathematical estimation less stable.

All three experiments included one or two control conditions to examine how the initial distractor stream impacted the TVA-parameters. As expected, the control condition in Experiment 1 revealed no significant difference in target processing before (Stream condition) and after the AB (Lag-8 condition). Further, Experiments 2 and 3 revealed a consistent pattern of results when comparing the control conditions in which the target display was presented with (Stream condition) and without (No-stream condition) the initial distractor stream. Here, we found a significantly lower perceptual threshold (lower $t_0$) and a significantly better selectivity (lower $\alpha$) when target displays were preceded by distractors. In TVA, the perceptual threshold, $t_0$, depends on the time taken to compute the attentional weights for targets and distractors in a given display. Fast computations lead to lower thresholds, whereas slow computations result in higher thresholds. Thus, we may speculate that the initial distractor stream facilitates the effectiveness of the template matching distinguishing distractors from targets in the target displays, resulting in a quick (lower $t_0$) and precise (lower $\alpha$) calculation of their attentional weights. Evidence for such a pre-target state in the AB paradigm has previously been reported by Dux et al. (2006) and by Maki and Padmanabhan (1994).

Overall, the three experiments consistently reproduce the same pattern of results. Most interestingly, they showed that during the AB, processing capacity, $C$, was significantly reduced, whereas the capacity of VSTM, $K$, and the efficiency of target selection, $\alpha$, were not significantly impaired. These results favor theories placing a bottleneck at the encoding stage, whereas less evidence is found in favor of
filter-based theories or theories placing a bottleneck at the maintenance stage. However, it remains an open question why the processing capacity of T2 is impaired during the AB. Here, the neural interpretation of TVA (NTVA; Bundesen et al., 2005) may provide further insight. Whereas TVA is a formal computational model, NTVA is a neurophysiological model using biologically plausible neural networks to implement TVA at the level of individual neurons. Based on NTVA, we suggest two hypotheses which can explain the observed impairment in processing capacity during the AB: 1) the number of cortical neurons available for T2 processing are fewer; or 2) the level of activation of the individual cortical neurons used for T2 processing is lower.

The theory of temporal visual attention (TTVA; Petersen et al., 2012) investigated the first hypothesis by extending NTVA to model data from the attentional dwell-time paradigm, a skeletal version of the AB paradigm in which only a post-masked T1 and T2 are presented on separate spatial locations without any distractors. TTVA assumes that T1 temporarily locks the cortical neurons used to encode T1 into VSTM until T1 has been recoded to a nonvisual (e.g., auditory, motoric, or amodel) format, leaving fewer cortical neurons available for T2 processing. The model convincingly fitted the classical U-shaped pattern observed by varying the time between T1 and T2 onsets, and also predicted data with a systematic variation in T2 exposure duration.

In a theoretical paper, Schneider (2013) extends on the ideas introduced by TTVA to not only explain the AB per se, but also other key findings from the AB literature (e.g., Lag-1 sparing). In his theory of task-driven visual attention and working memory (TRAM), Schneider suggests that the cortical neurons used to encode T1 into VSTM is only locked when a new competition episode calls for a clearance of VSTM and T1 needs to be protected for short-term consolidation (STC; Jolicoeur & Dell’Acqua, 1998). Schneider makes the assumption that a new competition episode is initiated when the attentional weight of the next item in the stream differs from the current item. According to TVA, distractors usually have lower attentional weights compared with targets whereas targets have equal weights. Thus, if T1 is followed by a distractor a new competition episode is initiated and T1 needs protection, but if T1 is followed directly by T2 no new competition episode needs to be initiated and Lag-1 sparing for T2 is observed as no locking of
cortical neurons is needed to protect T1. These additional assumptions of TRAM will, of course, have to be tested empirically in future studies.

Evidence for the second hypothesis comes from the *locus coeruleus norepinephrine* (LC-NE) model of the AB proposed by Nieuwenhuis et al. (2005). According to this neurocomputational model, the AB is the consequence of the dynamics of the LC responsible for the release of NE in the neocortex. At the level of individual neurons, the release of NE is believed to increase the responsiveness of neurons and have an attention-enhancing effect (Servan-Schreiber et al., 1990). It is hypothesized that the presentation of a target triggers LC activity which peaks around 100 ms after stimulus onset. After initial activation, LC enters a refractory period lasting approximately 400-450 ms (Aston-Jones et al., 1994) during which it is unable to enhance subsequent target processing. According to the LC-NE model, T1 triggers LC such that T1 can benefit from the NE-mediated enhancement. If T2 is presented during the following refractory period T2 will not receive this benefit, causing the AB. However, if T2 is presented at Lag 1 it will benefit from the residual NE-mediated enhancement elicited by LC response to T1 and be processed with greater accuracy, causing Lag-1 sparing.

Likewise, one of the most extensive computational models in the AB literature, the *episodic STST* (eSTST) model by Wyble et al. (2009), rely on excitation and inhibition of a transient attentional enhancement (TAE) resource, which, if activated, amplifies all input. In this extended version of the STST model, the presentation of a target triggers a blaster which boosts the encoding of T1 and all targets presented in succession after T1, explaining not only Lag-1 sparing but also spreading of the sparing (Kawahara et al., 2006; Olivers et al., 2007). However, an intervening distractor presented in between targets will provide sufficient time for the blaster to be suppressed, producing an AB for the following target. In the context of our findings, it is important to note that both the LC-NE model and the eSTST model are bottleneck theories in which the distractors in the RSVP stream and attentional control only play limited roles.

As for now, both hypotheses proposed on the basis of NTVA contribute to a potential explanation of the observed reduction in processing speed during the AB. To obtain evidence for one over the other
hypothesis, TVA-based research into other key findings in the AB literature (see review by Dux & Marois, 2009) is needed, but we consider this outside the scope of the current article.


