

Measuring and modeling attentional dwell time

Anders Petersen · Søren Kyllingsbæk · Claus Bundesen

Published online: 31 July 2012
© Psychonomic Society, Inc. 2012

Abstract Attentional dwell time (AD) defines our inability to perceive spatially separate events when they occur in rapid succession. In the standard AD paradigm, subjects should identify two target stimuli presented briefly at different peripheral locations with a varied stimulus onset asynchrony (SOA). The AD effect is seen as a long-lasting impediment in reporting the second target, culminating at SOAs of 200–500 ms. Here, we present the first quantitative computational model of the effect—a theory of temporal visual attention. The model is based on the neural theory of visual attention (Bundesen, Habekost, & Kyllingsbæk, *Psychological Review*, 112, 291–328 2005) and introduces the novel assumption that a stimulus retained in visual short-term memory takes up visual processing-resources used to encode stimuli into memory. Resources are thus locked and cannot process subsequent stimuli until the stimulus in memory has been recoded, which explains the long-lasting AD effect. The model is used to explain results from two experiments providing detailed individual data from both a standard AD paradigm and an extension with varied exposure duration of the target stimuli. Finally, we discuss new predictions by the model.

Keywords Attentional dwell time · Theory of visual attention · Attentional blink · Computational modeling

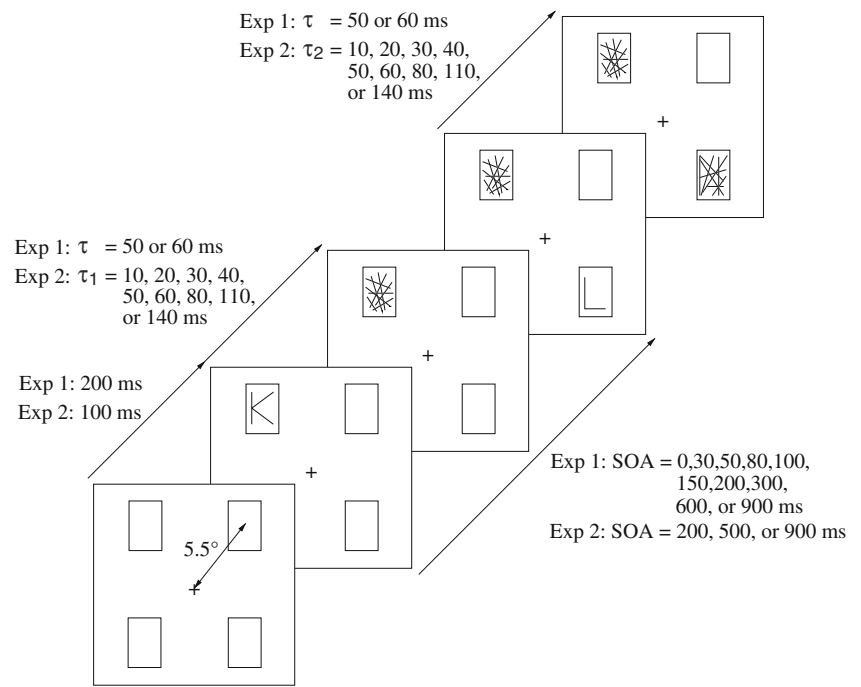
Our ability to allocate processing resources at different spatial locations across time is one of the core topics in visual attention research. Several different paradigms have been used to probe the nature of spatial shifts of visual

attention. In classic cuing paradigms, Posner and colleagues observed attentional shifts and called upon a metaphor that explains visual attention in terms of a spotlight highlighting a particular location in space (e.g., Müller & Rabbitt, 1989; Posner, 1980). Likewise, visual search paradigms have been used to probe the time course of spatial shifts in search for a target among distractors. Here, a main issue has been the extent to which attention operates in parallel versus serially across the visual field (e.g., Bundesen, 1990; Kyllingsbæk, Schneider, & Bundesen, 2001; Shiffrin & Gardner, 1972; Shiffrin & Schneider, 1977; Treisman & Gelade, 1980; Wolfe, 1994). However, in many of these paradigms, the detailed time course of individual shifts of attention has been elusive. In visual search paradigms, for example, the measured reaction times may include several shifts of attention between stimuli or groups of stimuli, obscuring the time course of individual shifts of attention. Thus, to study individual shifts of attention, a simpler paradigm is needed.

Duncan, Ward, and Shapiro (1994; see also Ward, Duncan, & Shapiro, 1996) proposed the *attentional dwell time* (AD) paradigm as a simpler alternative when the time course of visual attention is investigated. In the standard AD paradigm, two target stimuli (T1 and T2) are presented at peripheral locations around a central fixation cross (see Fig. 1). Presentations are brief—typically, around 50 ms—and the target stimuli are followed by pattern masks to prevent further processing after their offset. The stimulus onset asynchrony (SOA) is varied systematically from 0 to around 1,000 ms, and subjects are instructed to make an unspeeded report of the identity of the targets. The AD effect is seen as an impediment in reports of T2 culminating at onset-to-onset times of 200–500 ms. Furthermore, the effect is surprisingly long-lasting; thus, report of T2 is independent of presentation of T1 only after 1 s has passed (see Fig. 2).

A. Petersen (✉) · S. Kyllingsbæk · C. Bundesen
Department of Psychology, University of Copenhagen,
Copenhagen, Denmark
e-mail: anders.petersen@psy.ku.dk

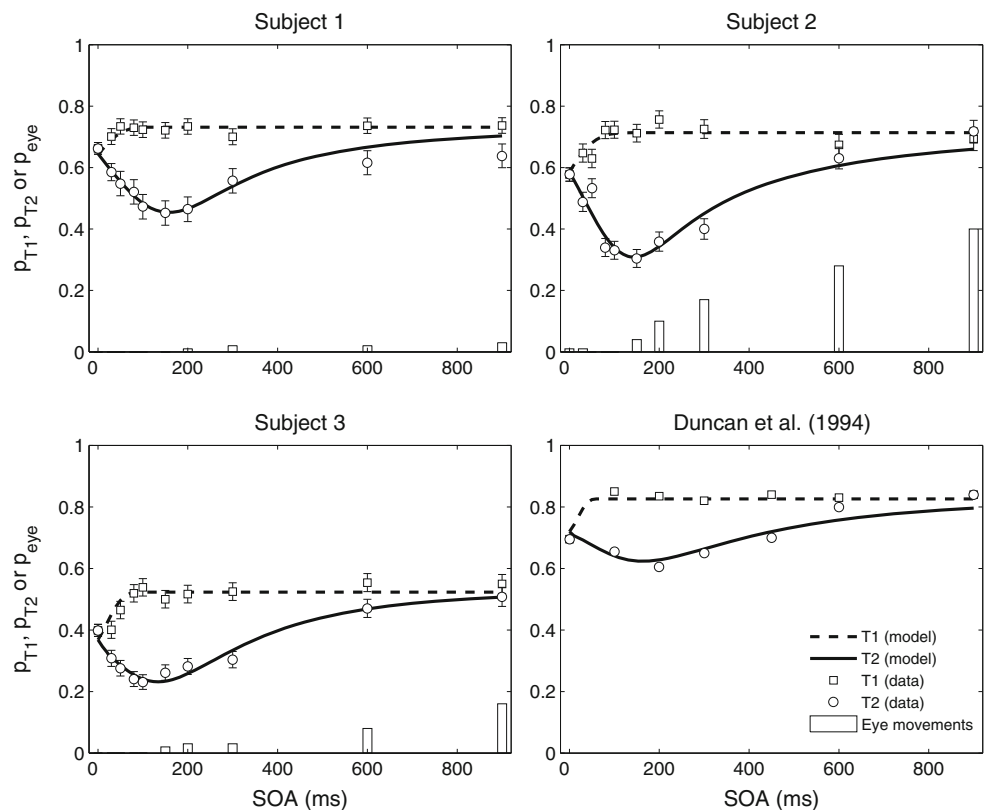
Fig. 1 Experimental setups. A short fixed delay preceded the presentations of T1, while the SOA between T1 and T2 was varied. The exposure durations of T1 and T2 were fixed in Experiment 1 but varied in Experiment 2



Ward, Duncan, and Shapiro (1997) related findings from the AD paradigm to a simplified version of the *attentional blink* (AB) paradigm where T1 and T2 have to be identified in a centrally presented stream of distractors (rapid visual serial presentation [RSVP]; e.g., Broadbent & Broadbent,

1987; Chun & Potter, 1995; Potter & Levy 1969; Raymond, Shapiro, & Arnell, 1992). In their simplified version, T1 and T2 were backward masked and presented on the same spatial location, but without the stream of distractors. The time course of effects seen in this simplified version of the

Fig. 2 Results of Experiment 1: The probabilities of correctly reporting T1 (p_{T1} , squares) and T2 (p_{T2} , circles) as functions of SOA for each of the 3 subjects in Experiment 1 and for the group average reported in Duncan et al. (1994). The exposure duration of both targets was $\tau = 60$ ms for subjects 1 and 3 and $\tau = 50$ ms for subject 2. In Duncan et al., the average exposure duration was $\tau_{\text{average}} = 57$ ms. The plotted standard deviations are calculated assuming that the responses from the subjects were approximately binomially distributed. Furthermore, the probability of making an eye movement (p_{eye} , bars) is displayed as a function of SOA for each of the 3 subjects in Experiment 1. Finally, the least squares fits of the TTVA model to each of the four data sets are plotted for T1 (dashed line) and T2 (solid line)



AB paradigm was similar to the time course observed in the AD paradigm, and thus Ward et al. (1997) argued for a common underlying mechanism. In this article, we have chosen to focus on the AD paradigm by Duncan, Ward, and Shapiro (1994). This decision was made because we wanted the simplest possible setup to measure and model the temporal dynamics of attention. Using the AD paradigm, we do not have to deal with distractors as presented in the AB paradigm or the undesired premasking of T2 by T1 that occurs in the paradigm by Ward et al. (1997) when T2 is presented in close temporal proximity to T1.

Although results from the initial studies using the AD paradigm attracted much attention, only a few subsequent studies using the paradigm have been reported. In two such studies, the effect of the masks following T1 and T2 was investigated: Moore, Egeth, Berglan, and Luck (1996) found that the duration of the AD effect was reduced significantly to about 200 ms when presentation of the mask for T1 was postponed until the offset of T2. Brehaut, Enns, and Di Lollo (1999) extended these results by investigating the effect of using an integration mask versus an interruption mask. The AD effect was found only when T2 was masked by interruption. Recently, Petersen and Kyllingsbæk (in press) presented an extensive study of the effect of eye movements and practice in the AD paradigm. In the previous studies, each subject ran fewer than 1,000 trials, and Ward et al. (1996) reported little or no effect of practice across the 2 days the subjects were tested. In contrast, Petersen and Kyllingsbæk found a strong reduction in the AD effect following 6 days of intensive practice corresponding to a total of 4,680 trials. For some of the subjects, the AD effect was virtually absent on the final day of testing. It was found that controlling for eye movements and using masks that varied from trial to trial, rather than a fixed mask as used in the classical AD experiments, counteracted the effect of practice and led to a stable AD effect across the six test sessions.

In this article, we present results from two experiments providing detailed individual data from both a traditional AD paradigm and an extension where we varied the exposure duration of the two target stimuli. To explain the results of the experiments, we propose the first computational model of the AD effect. The model is based on the neural theory of visual attention (NTVA) by Bundesen, Habekost, and Kyllingsbæk (2005; see also Bundesen, 1990) and introduces the novel assumption that retention of a stimulus (e.g., T1) to be remembered in visual short-term memory (VSTM) takes up visual-processing resources used to identify the stimulus. Until the stimulus is recoded into a nonvisual (e.g., auditory, motoric, or amodal) format, the resources are locked and cannot be used to encode subsequent stimuli (e.g., T2) into VSTM. This mechanism creates a temporary encoding bottleneck that explains the time course of the AD. Other computational models of temporal attention have

primarily focused on the AB, but none of these models have accounted for data from the AD paradigm.

Experiment 1

In Experiment 1, accurate measures of the time course of the AD in individual subjects were made, requiring subjects to be tested for a considerable number of trials to minimize noise. Such accurate measures were necessary in order to thoroughly test the proposed model. We incorporated the improvements suggested by Petersen and Kyllingsbæk (in press) into the experiment. Thus, we analyzed only trials without eye movements and varied the mask from trial to trial. This enabled us to map the time course of the AD with great precision.

Method

Subjects Three psychology students (all female; mean age = 27 years) from the University of Copenhagen were paid a standard fee by the hour for participating in the experiment. All had normal or corrected-to-normal vision.

Targets The targets were all 26 uppercase letters of the English alphabet constructed from 27 unique line segments. The letters were white, with a width and height of 1.10° and 1.65° , respectively.

Masks The masks varied from trial to trial in order to ensure that subjects did not habituate to the pattern of any particular mask. The masks were constructed from the same 27 unique line segments that were used for constructing the stimulus letters (see Fig. 1). Each mask was made by randomly choosing 14 of the 27 unique line segments and shifting the 14 segments independently of each other 0.55° to the left (probability .2), 0.55° to the right (probability .2), 0.55° up (probability .2), 0.55° down (probability .2), or not at all (probability .2). This procedure made the size of the masks slightly larger than the size of the letters.

Procedure Stimuli were presented on a 19'-in. CRT monitor at 100 Hz, using in-house custom-made software written in C++. A white fixation cross ($0.55^\circ \times 0.55^\circ$) was displayed on a black background together with four white boxes serving as place holders ($1.65^\circ \times 2.20^\circ$). The boxes were placed at the corners of an imaginary square 5.5° from fixation (see Fig. 1). Subjects initiated every trial themselves by pressing the space bar on the keyboard. After a delay of 200 ms, the first target letter (T1) appeared in one of the boxes, followed by a mask that stayed on the screen until the end of the trial. After an SOA of 0, 30, 50, 80, 100, 150, 200, 300, 600, or 900 ms, a second target (T2) was presented in one of the

three remaining boxes, followed by a second mask for 240 ms. The subjects responded by typing the letters on a keyboard in any order preferred. A forced choice procedure was applied; that is, subjects should respond to each of the targets even if they had to guess.

Subjects were instructed to maintain central fixation during trials but were allowed to move their eyes between trials. Eye movements were measured using a head-mounted eye tracker (EyeLink II). A trial was categorized as a *trial with eye movements* if the gaze deviated more than 2.75° from the fixation cross (i.e., half the distance from the fixation cross to the target locations) at any time during the trial before the onset of the second mask. We did not measure eye movements after the onset of the second mask, assuming that they would not influence the identification of T2. If no eye movements were registered outside the 2.75° boundary, the trial was categorized as a *trial without eye movements*.

All subjects did six sessions on 6 different days. Each session comprised 50 practice trials and two blocks of experimental trials. Before the first block of the first session, subjects performed a calibration procedure to avoid floor or ceiling effects. The calibration used an adaptive psychophysical procedure (accelerated stochastic approximation; Kesten, 1958) that adjusted the exposure duration of the letters such that, on average, one of the two letters (i.e., about 50 % of the presented letters) were correctly reported in the condition in which the two letters were presented simultaneously (i.e., $SOA = 0$). The procedure comprised 50 trials on which the exposure duration on the current trial was adjusted on the basis of the exposure duration and the number of correctly reported targets on the previous trial. Each subject performed three calibrations. The first calibration was initiated using an exposure duration of 80 ms, and the following two calibrations used the outcome of the previous calibration as the starting point. The calibration resulted in 60-ms exposure duration for subjects 1 and 3 and 50-ms exposure duration for subject 2.

Design Letters were chosen randomly without replacement so that each letter was used once and only once as the first and second targets for each SOA, to avoid variation caused by different salience of the letters. Furthermore, letters were chosen so that within a trial, the two target letters were always different. With 10 SOAs and 26 letters, one block of the experiment consisted of 260 trials. Thus, the entire experiment comprised 3,120 experimental trials. Target locations were pseudorandomized such that within one block of the experiment, T1 and T2 were presented equally often in all four boxes.

Results and discussion

Figure 2 shows the mean probability of correctly reporting T1 (p_{T1}) and T2 (p_{T2}) as a function of SOA for each of the 3

subjects. For the condition in which both targets were presented simultaneously (i.e., $SOA = 0$ ms), the average of p_{T1} and p_{T2} is plotted.

Only trials without eye movements were included in the above calculations. Although subjects were instructed to maintain central fixation, they nevertheless made eye movements. According to Petersen and Kyllingsbæk (2012), trials with eye movements should be excluded from the data analysis because they confound the AD effect. The bars in Fig. 2 show the proportion of trials with eye movements (p_{eye}) as a function of SOA for each subject.

Finally, Fig. 2 shows the standard deviations of p_{T1} and p_{T2} . The standard deviations were calculated assuming that responses from the subjects were binomially distributed; that is, we assumed stochastic independence of the trials and a constant probability p of correctly reporting a target. The standard deviation for a probability p is then given by $SD = \sqrt{p(1-p)/N}$, where N is the number of trials in the condition.

Letter identification All 3 subjects showed a fast gradual decrease in p_{T2} as the SOA increased from 0 to 100–200 ms, followed by a slow improvement as the SOA increased beyond 200 ms. Duncan et al. (1994) found a similar time course of p_{T2} (cf. Fig. 2) and reported that p_{T1} was higher when T1 was presented alone (i.e., $SOA > \tau$), as compared with when it was presented together with T2 (i.e., $SOA = 0$). All 3 subjects showed a corresponding increase in p_{T1} . Furthermore, our data suggest that the increase occurred gradually.

The data also suggest some individual variation in letter identification. The most noticeable difference was observed in the magnitude of the impairment of p_{T2} , with subject 2 producing a notably larger impairment than did subjects 1 and 3.

Eye movements Individual variation was also found in the proportion of trials with eye movements. Subject 1 made very few eye movements, whereas a higher proportion of trials with eye movements was recorded for subjects 2 and 3—in particular, at the longer SOAs. Eye movements were measured until the onset of the second mask, which suggests that the recorded eye movements were prompted by the presentation of T1. This may explain why more eye movements were recorded at long SOAs, as compared with short SOAs: If T2 was presented in close temporal proximity to T1, programming of an eye movement toward T1 was interrupted by the onset of T2. Thus, the longer the SOA, the smaller the likelihood that the onset of T2 interrupted the execution of an eye movement toward T1.

In summary, Experiment 1 replicated the AD effect at the level of the individual subjects. The measurements of the

time course of the AD turned out to be very accurate and revealed individual differences especially in the magnitude of the AD.

Experiment 2

In Experiment 2, we investigated performance in the AD paradigm by varying both the SOA between T1 and T2 and the exposure duration of the letters. This added an additional unexplored dimension to the data and provided new important information about the AD effect.

Investigating attentional effects by altering the exposure duration of stimuli is not a new idea. Sperling (1967) systematically varied the exposure duration of the targets in his whole-report paradigm and found an increase in the number of reported letters as the exposure duration increased. Sperling argued that this was evidence not only of limited storage capacity of VSTM, but also of a limitation in how fast letters can be encoded into VSTM. Systematic variation of exposure duration was also used in the partial-report paradigm by Shibuya and Bundesen (1988), with results suggesting that the process of encoding the targets and distractors had a fixed capacity limitation. This finding among others later led to the development of Bundesen's (1990) theory of visual attention (TVA).

However, when it comes to investigations of dual tasks like the AB paradigm, only a few studies have systematically altered the exposure duration of targets (Jolicoeur & Dell'Acqua, 1999, 2000; McLaughlin, Shore, & Klein, 2001). To our knowledge, this is the first time that exposure duration of targets has been varied in the AD paradigm. From a modeling perspective, we find this to be a crucial manipulation if one wants to understand the mechanisms behind the AD effect. In the special case in which T1 and T2 are presented simultaneously (i.e., whole report), TVA has already provided a detailed mathematical description of how an increase in exposure duration leads to more correctly reported targets. Experiment 2 should show us if essentially the same description can be used when T1 and T2 are presented with a temporal gap.

Method

Subjects Three psychology students (all female; mean age = 21.7 years) from the University of Copenhagen participated in the experiment and were paid a standard fee by the hour. All had normal or corrected-to-normal vision.

Targets The targets were all 26 uppercase letters presented in Elektra font (available on a free license from <http://www.dafont.com>). The letters in the Elektra font are made

up of small black boxes placed on a grid containing a total of five boxes in the horizontal direction and seven boxes in the vertical direction. Our letters had a width and a height of 0.65° and 0.98° , respectively.

Masks Masks were constructed by randomly placing black boxes in half of the locations in a grid containing seven boxes in the horizontal direction and nine boxes in the vertical direction. Thus, the masks were slightly bigger than the letters, with a width of 0.90° and a height of 1.23° . In total, 26 masks were constructed using this procedure. However, if a mask was constructed with a very uneven distribution of black boxes, it was replaced by a new mask in an effort to ensure that all masks would be equally efficient.

Procedure Stimuli were presented on a 19'-in. CRT monitor at 100 Hz using E-Prime 2.0 software. A black fixation cross ($0.41^\circ \times 0.41^\circ$) was displayed on a gray background. Subjects initiated a trial by pressing the space bar, and after a delay of 100 ms, the first target letter (T1) appeared at one of four possible locations. Similar to Experiment 1, target locations were at the corners of an imaginary square 3.5° from the fixation cross, but in contrast to Experiment 1, the locations were not marked by place holders. T1 was presented for τ_1 ms, followed by a mask that was presented until the end of the trial. Afterward, the second target (T2) was presented at one of the three remaining locations for a duration of τ_2 and then masked for 240 ms. The SOA between the onsets of T1 and T2 was 200, 500, or 900 ms (see Fig. 1). As in Experiment 1, subjects responded by typing the letters on a keyboard in any order preferred and were required to always report two letters (i.e., a forced choice procedure).

In contrast to Experiment 1, exposure durations were varied systematically in this experiment. Thus, no initial calibration of exposure duration was done. Subjects performed only a short practice block of 25 trials before starting each experimental block. Subjects ran four blocks of the experiment on 4 different days.

Identical to the procedure of Experiment 1, subjects were required to maintain central fixation during trials but were allowed to move their eyes between trials. A table-mounted eye tracker (Eyelink 1000) was used to ensure that no eye movements were made during trials. Similar to Experiment 1, a trial was categorized as a *trial with eye movements* if the gaze moved more than 1.75° away from the fixation cross (i.e., half the distance from the fixation cross to the target locations). In contrast to Experiment 1, trials with eye movements were reinserted and rerun at the end of a block, ensuring a full data set regardless of the frequency of eye movements.

Design The experiment comprised four conditions. In condition 1, the exposure duration of T1 (τ_1) was varied while the exposure duration of T2 was kept constant ($\tau_2 = 80$ ms). In conditions 2–4, the exposure duration of T2 (τ_2) was varied while the exposure duration of T1 was constant ($\tau_1 = 80$ ms). The SOA in condition 1 was 900 ms, whereas the SOA in conditions 2, 3, and 4 was 200, 500, and 900 ms, respectively (see Table 1). In each condition, nine different exposure durations were used: 10, 20, 30, 40, 50, 60, 80, 110, and 140 ms. As in Experiment 1, stimulus letters were chosen randomly without replacement, such that each letter was used once and only once as the first and second targets in every combination of condition type and exposure duration. Furthermore, the two letters within a trial had to be different. We used a factorial design where condition type, exposure duration, and letter type were randomly intermixed. Thus, one block of the experiment comprised 936 trials (4 conditions \times 9 exposure durations \times 26 letter types), and the entire experiment comprised 3,744 experimental trials, excluding trials with eye movements. In contrast to Experiment 1, target locations were selected at random without any constraints (i.e., without any balancing).

Results and discussion

Figures 3 and 4 show the results of Experiment 2. Figure 3 shows (a) the proportion p_{T2} of correct reports of T2 as a function of exposure duration τ_1 of T1 when τ_1 was varied (i.e., in condition 1) and (b) the proportions p_{T1} of correct reports of T1 as functions of exposure duration τ_2 of T2 when τ_2 was varied (i.e., in conditions 2–4). The flat lines are the model predictions. Figure 4 shows (a) the proportion p_{T1} of correct reports of T1 as a function of its exposure duration in condition 1 and (b) the proportions p_{T2} of correct reports of T2 as functions of its exposure duration in conditions 2–4. The sigmoid lines are the model predictions. In all conditions, the probability of correctly reporting a letter with an exposure duration of 10 ms was at the level of blind guessing (1/26). As the exposure was increased from 10 to 140 ms, the probability of correctly reporting the letter increased as a sigmoid function of the exposure duration. The highest rates of increase were found in conditions 1 and 4, in which the longest SOA (900 ms) was used. In these conditions, the processing of T2 seemed unaffected by the presentation of T1, and vice versa. By contrast, if T1 and T2 were presented in close temporal proximity (SOA = 200 ms, condition 2), the presentation of T1 strongly reduced the rate of increase in the probability of correctly reporting T2 as a function of its exposure duration (see the lowest curve in each of the three panels in Fig. 4). This reduction was gradually attenuated as the SOA was increased to 500 ms (condition 3) and further increased to 900 ms (condition 4). Thus, similar to Experiment 1, a transient impairment in

correctly reporting T2 was found in Experiment 2. In summary, both Experiments 1 and 2 yielded highly systematic dwell time data for individual subjects. In the remaining part of this article, we will focus on the development of a mathematical model to account for these data.

A theory of visual attention

The theoretical basis of the model is the TVA (Bundesen, 1990). TVA is a computational model, which makes it possible to make quantitative predictions of results from experiments on visual attention. This section gives a short introduction to TVA and a neural interpretation of TVA (NTVA), whereas the following section describes the principles behind a temporal extension of TVA accounting for the AD data.

TVA has accounted for a wide range of experimental findings, including effects of object integrality (Duncan, 1984), varying numbers of targets in studies of divided attention (Sperling, 1960, 1967), varying numbers of targets and distractors in partial report (Bundesen, Pedersen, & Larsen, 1984; Bundesen, Shibuya, & Larsen, 1985; Shibuya & Bundesen, 1988), selection criterion and set size in visual search (Treisman & Gelade, 1980), and practice in visual search (Schneider & Fisk, 1982). TVA proposes that visual recognition and attentional selection of elements in the visual field consist in making perceptual categorizations. A perceptual categorization has the form "object x belongs to category i " (or equivalently "object x has feature i "), where x is an object in the visual field and i is a perceptual category (e.g., a certain color, shape, movement, or spatial position). An object x is said to be encoded into VSTM if a categorization of the object is encoded into VSTM. The categorizations are assumed to be processed mutually independently and in parallel. The "speed" at which a particular categorization is encoded into VSTM is determined by the hazard function, $v(x,i)$ —that is, the density function of the conditional probability that a categorization will occur at time t , given that the categorization has not occurred before time t . Parameter $v(x,i)$ is also referred to as the *processing rate* of categorizing object x as having feature i and is calculated using the *rate equation* of TVA,

$$v(x,i) = \eta(x,i)\beta_i \frac{w_x}{\sum_{z \in S} w_z} \quad (1)$$

where $\eta(x,i) \in \mathbb{R}_+ \cup \{0\}$ is the strength of the sensory evidence that object x belongs to category i , $\beta_i \in [0,1]$ is the perceptual bias associated with category i , and $w_x \in \mathbb{R}_+ \cup \{0\}$ is the attentional weight of object x which is divided by the sum of attentional weights across all objects in the visual field, S .

In many applications of TVA, it is convenient to define the total processing rate of an object, v_x , as the sum of the

processing rates of all categorizations of object x —that is,

$$v_x = \sum_{i \in R} v(x, i) = \sum_{i \in R} \eta(x, i) \beta_i \frac{w_x}{\sum_{z \in S} w_z} = s_x \frac{w_x}{\sum_{z \in S} w_z}, \quad (2)$$

where R is the set of all perceptual categories and $s_x = \sum_{i \in R} \eta(x, i) \beta_i$ is referred to as the *sensory effectiveness* of object x . Furthermore, it is convenient to define the *total processing capacity*, C , as the sum of processing rates across all perceptual categories, R , and all elements in the visual field, S —that is,

$$C = \sum_{x \in S} \sum_{i \in R} v(x, i) = \sum_{x \in S} v_x \quad (3)$$

Thus, if all objects presented in a display have the same sensory effectiveness s (homogeneous display), then by inserting Eq. 2 into Eq. 3, we find that

$$C = \sum_{x \in S} s \frac{w_x}{\sum_{z \in S} w_z} = s \quad (4)$$

We may postulate that the sensory effectiveness is the same for all letters in a paradigm if we make the following assumptions: Assume that the bias values for all letter types, $\beta_A, \beta_B, \dots, \beta_Z$ equal a positive constant β but, for any other perceptual categories (e.g., color, size, etc.) the bias parameters equal zero. Then the sensory effectiveness can be simplified to $s_x = \sum_{i \in \{A, B, \dots, Z\}} \eta(x, i) \beta$. Furthermore, assume that for any letter x different from letter type i , $\eta(x, i) = 0$ (i.e., perceptual confusion errors are neglected) but, for any letter x of letter type i , $\eta(x, i)$ equals a positive constant η for all $i \in \{A, B, \dots, Z\}$. Then, the sensory effectiveness reduces to $s = \eta \beta$ and is the same for all letters.

In this special case of TVA, Eq. 4 can be inserted into Eq. 2, and the processing rate of a letter x can be calculated by the simplified equation,

$$v_x = C \frac{w_x}{\sum_{z \in S} w_z} \quad (5)$$

where C is now the total processing capacity for letters. This simplification leads to a *fixed-capacity independent race model* (Shibuya & Bundesen, 1988) in which the visual system is assumed to have a limited processing capacity and objects in the visual field compete for these limited processing resources. The competition is represented by the attribution of attentional weights. Objects with high weights will get more processing resources than will objects with low weights. Objects will then *race* against each other to access VSTM. The more processing resources allocated to an object, the higher is the probability that the object will be encoded into VSTM. Only objects encoded into VSTM can be reported correctly without guessing. In TVA, VSTM is assumed to have a limitation as to how many object it can hold at any given time. For normal subjects, the capacity of VSTM (K) is around three to four objects.

If the number of presented objects does not exceed K and stimulus processing is interrupted by a mask presented at stimulus offset, the probability of encoding an object into VSTM has traditionally in TVA been given by

$$p = 1 - e^{-v_x(\tau - t_0)} \text{ for } \tau > t_0 \quad (6)$$

where v_x is the *processing rate* of object x , τ is the exposure duration of object x , and t_0 is the longest ineffective exposure duration (a.k.a. the threshold for visual perception). That is, if the exposure duration of an object is shorter than t_0 , the probability of encoding the object into VSTM will be zero. However, if the exposure duration of an object is longer than t_0 , the processing time of the object is assumed to be exponentially distributed, resulting in an exponential increase in the probability of encoding the object into VSTM as a function of exposure duration.

Recently, Dyrholm, Kyllingsbæk, Espeseth, and Bundesen (2011) have provided evidence suggesting that some variation in t_0 across trials must be incorporated into TVA. This may be achieved by assuming that t_0 is approximately normally distributed with mean μ_0 and standard deviation σ_0 . When limitations of the storage capacity of VSTM can be neglected, the probability of encoding an object into VSTM is then approximated by

$$p = \int_{-\infty}^{\tau} \frac{1}{\sigma_0} \phi\left(\frac{t_0 - \mu_0}{\sigma_0}\right) \left(1 - e^{-v_x(\tau - t_0)}\right) dt_0 \\ = \Phi\left(\frac{\tau - \mu_0}{\sigma_0}\right) - e^{-v_x(\tau - \mu_0 - \frac{1}{2}\sigma_0^2 v_x)} \Phi\left(\frac{\tau - \mu_0 - \sigma_0^2 v_x}{\sigma_0}\right) \quad (7)$$

where $\frac{1}{\sigma_0} \phi\left(\frac{t_0 - \mu_0}{\sigma_0}\right)$ is the probability density function of t_0 if t_0 is normally distributed with mean μ_0 and standard deviation σ_0 . $\phi(x)$ and $\Phi(x)$ are the probability density function and the cumulative distribution function of the standard normal distribution, respectively. In other words, the time it takes for an object to be encoded into VSTM is modeled as coming from an ex-Gaussian distribution (i.e., a convolution of a normal distribution and an exponential distribution; Luce, 1986).

The neural interpretation of TVA

The neural interpretation of TVA (NTVA; Bundesen et al., 2005) is a further development of TVA. Whereas TVA is a formal computational theory, NTVA is a neurophysiological model using biologically plausible neural networks to implement the equations of TVA at the level of individual neurons. In NTVA, the number of cortical neurons representing an object x is proportional to the relative attentional weight of the object, $w_x / \sum_{z \in S} w_z$, and the level of activity in the neurons representing object x corresponds to the

multiplicative scaling of the relative attentional weight by s_x (or C if the display is homogeneous). Each neuron is regarded as representing the properties of only one object at a time, which is supported by evidence from single-cell studies (Moran & Desimone, 1985). In this way, cortical neurons are distributed among the objects in the visual field so that each object will race toward VSTM with an individual processing rate $v_x = s_x w_x / \sum_{z \in S} w_z$.

The implementation of VSTM in NTVA builds on the Hebbian notion that short-term memory is based on retaining information in feedback loops that sustain activity in the neurons representing the information (Hebb, 1949). In NTVA, a categorization of an object x becomes encoded in VSTM by becoming embedded in a positive feedback loop—a feedback loop, which is closed when, and only when, a unit representing object x in a topographic map of objects (the *VSTM map*) is activated. Thus, impulses routed to a unit that represents an object at a certain location in the topographic VSTM map of objects are fed back to the feature units from which they originated, provided that the VSTM unit is activated. If the VSTM unit is inactive, impulses to the unit are not fed back. Thus, for each feature- i neuron representing object x , activation of the neuron is sustained by feedback when the unit representing object x in the topographic VSTM map of objects is activated.

The storage limitation of K objects in VSTM is implemented as a K -winners-take-all-network in which all nodes have inhibitory connections to all other nodes in the network and excitatory connections (feedback loops) to themselves. When fewer than K objects are encoded, the inhibitory activation within the network is low enough to allow activation of additional nodes (encoding of more objects). However, if K objects have been encoded, the inhibitory activation within the network is so high that activation of additional nodes will not be possible.

A theory of temporal visual attention

TVA has been applied to a wide range of behavioral paradigms. However, most of these paradigms have used simultaneous presentation of stimuli (e.g., whole report and partial report). The AD paradigm introduces a temporal dimension that has not previously been accounted for by TVA or NTVA. In this section, we will introduce a *theory of temporal visual attention* (TTVA), which aims to explain how states of attention change over time. The model is a generalization of TVA and reduces to TVA in the special case in which stimuli are presented simultaneously.

Multiple races for encoding

When TVA is applied to paradigms with simultaneous presentation of letters, it is assumed that only one calculation of

attentional weights is performed, followed by a single race among the letters to become encoded into VSTM. Following Shibuya and Bundesen (1988), let t_1 be the time from the stimulus display is presented until the race toward VSTM is initiated, let t_2 be the time from the postmask is presented until the race is interrupted by the mask, and let t_0 be the difference between t_1 and t_2 —that is, $t_0 = t_1 - t_2$. Then the race lasts for a time equal to the stimulus duration τ minus t_0 , provided $\tau > t_0$. If $\tau \leq t_0$, no race is run. Thus, parameter t_0 is the longest ineffective stimulus duration.

The probability that the stimulus becomes encoded is a function of $\tau - t_0$ (cf. Eq. 6), but independent of t_1 and t_2 as long as the difference between the two times, t_0 , is kept constant. Thus, our model predictions would be the same if t_1 equaled t_0 , while t_2 was 0. For ease of exposition, we shall suppose that calculation of attentional weights begins when the letters are presented (time 0) and finishes such that the race based on the weights can begin at time t_0 . The race is interrupted as soon as the mask is presented, at time τ , yielding a race duration of $\tau - t_0$.

In temporal paradigms such as the AD paradigm, an extended way of thinking about the calculation of attentional weights and the following race is required because letters are no longer presented simultaneously (except at SOA = 0). The simplest extension is to introduce multiple calculations of attentional weights such that attentional weights are redistributed and a new race is initiated every time a letter is presented. Thus, when two letters are presented with a temporal gap, two calculations of attentional weights will be initiated—one when T1 is presented and one when T2 is presented. This implies that the calculations will finish at separate time points, resulting in two redistributions of the attentional resources—one at t_{01} (i.e., t_0 for T1) and one at t_{02} (i.e., t_0 for T2). Following each redistribution, a new race toward VSTM will be initiated. For both T1 and T2, the race lasts until the masks destroy their representations—that is, until time τ_1 for T1 and until time SOA + τ_2 for T2.

Locking of resources

As previously described, NTVA introduces a feedback mechanism from the VSTM map of objects back to the visual-processing neurons representing the encoded features of the objects. This sustains activity in the neurons and, thus, the representation of an encoded object in VSTM—a feedback loop. In paradigms such as whole and partial report, a feedback loop can sustain either the representation of an encoded letter or some representation of the subsequent mask—the latter if the letter was not encoded before the mask destroyed its representation. In this case, subjects may categorize part of the mask as a letter and sustain this representation in a feedback loop.

In a scenario with multiple races, a similar feedback mechanism may exist if we assume that neurons already engaged in a feedback loop are prevented from being reallocated to process other objects in a later race without first being disengaged from the loop. We say that neurons are *locked* in a feedback loop to retain a representation of the encoded object in VSTM. In TTVA, the time it takes to lock a neuron in a feedback loop is assumed to be exponentially distributed with a rate parameter λ_l and is referred to as the *lock-time*. The exponential distribution may also be described by its mean $\mu_l = 1/\lambda_l$ which we will refer to as the *mean lock-time*.

Release of resources

Inspired by the *attentional dwell-time hypothesis* (Ward et al., 1996), we propose that after a neuron has been locked to retain a representation of an encoded object in VSTM, it will dwell on the object such that the representation can be recoded into a more permanent storage for later report. We refer to this as the *dwell-time* of the locked neurons. As for the lock-time, we assume that the dwell-time is exponentially distributed with a rate parameter λ_d and refer to the mean of the distribution as the *mean dwell-time* ($\mu_d = 1/\lambda_d$). After the dwell-time has passed, a neuron is *released* (i.e., the feedback loop is broken) and can be redistributed to process other objects. However, one may assume that even though the loop is broken, feedback from VSTM is still active and guides the neuron such that it is not redistributed to process the same object. This seems a plausible assumption, since a similar guiding mechanism has been reported in studies of *inhibition of return* (Posner & Cohen, 1984; Klein, 2000). These studies have found that after attention is removed from a previously attended peripheral location, there is a delay in responding to subsequent stimuli displayed at the same location.

Modeling the attentional dwell time effect

The locking and releasing of visual-processing neurons can be combined to account for the attentional dwell time effect observed in Experiments 1 and 2. When T1 and T2 are presented simultaneously (i.e., SOA = 0), t_{01} and t_{02} follow the same normal distribution with a mean μ_0 and a standard deviation σ_0 . This reduces TTVA to the traditional TVA model with only one race initiated at time t_0 . In this special case, neurons will be distributed equally among the two letters, since the relative attentional weight for either letter equals $\frac{1}{2}$. This implies that the processing rate of either letter is $\frac{1}{2}C$, and we predict equal probabilities of encoding T1 and T2.

If, however, T1 has a head start (i.e., SOA > 0), t_{01} and t_{02} will no longer follow the same normal distribution: t_{01} will be

normally distributed with mean μ_0 and standard deviation σ_0 whereas t_{02} will be normally distributed with mean SOA + μ_0 and standard deviation σ_0 . Thus, when T1 has a head start, it is most likely that $t_{01} \leq t_{02}$. In the interval between t_{01} and t_{02} , only the attentional weight of T1 will be available. In this interval, T1 will have an advantage over T2, since all neurons will be allocated to process T1 such that the processing rate of T1 is C . The head start of T1 results in a further advantage as neurons processing T1 (or the mask for T1) in the interval between t_{01} and t_{02} may be locked to retain the representation of T1 in VSTM. Consequently, T2 will lack processing resources when it is presented approximately 100–200 ms after T1 (see Fig. 2).

However, T1 loses this advantage as the interval between t_{01} and t_{02} increases, because more neurons will be released from T1 and become available for the processing of T2. Thus, when the interval between t_{01} and t_{02} is long, all neurons will have been locked to and released from T1 and are now exclusively available for the processing of T2. Thus, when T2 is presented approximately 900 ms after T1, both letters will have a processing rate of C .

The model sketched above explains the temporal impairment in correctly reporting T2 observed in the AD paradigm. The model also explains the improvement in correctly reporting T1 when T1 is presented alone (i.e., SOA > τ_1) and the equal probability of correctly reporting T1 and T2 when the two letters are presented simultaneously (i.e., SOA = 0). For a more formal description of the model, see the [Appendix](#).

Fits

The presented model has five free parameters: μ_l (mean lock-time), μ_d (mean dwell-time), C (total processing capacity), μ_0 (mean of the longest ineffective exposure duration, t_0), and σ_0 (standard deviation of t_0). A least-squares method was used to fit the model to the behavioral data in Experiments 1 and 2. Furthermore, to show that the model can also be used to explain existing AD data, it was fitted to the data from Duncan et al. (1994). In all three experiments, the subjects were forced to respond—if necessary, by guessing. To account for the guessing, we used a high-threshold guessing model, which assumes that a subject reports the identity of a target correctly if the target becomes encoded into VSTM¹ but, if the target fails to become encoded into VSTM, the subject guesses at random among the N alternatives. Formally, the

¹ The assumption implies that a target that becomes encoded into VSTM is retained or recoded so well that the identity of the target can be reported even if a second target with the same attentional weight as the first target competes with the first one for processing resources. This simplification seems plausible in view of our presumption that up to K independent items can be retained in VSTM without noticeable interference between them.

adjusted probability of correctly reporting a target T using this guessing model can be defined as

$$p_T^{adj}(\text{SOA}) = p_T(\text{SOA}) + (1 - p_T(\text{SOA})) \frac{1}{N} \quad (8)$$

In the experiment by Duncan et al. (1994), only two letter types for T1 and two digit types for T2 were used. Therefore, we used $N = 2$ when the model was fitted to these data. In Experiments 1 and 2, 26 different letter types were used for both T1 and T2. Thus, we used $N = 26$ when the model was fitted to the data in Experiments 1 and 2.

Figure 2 shows the model fit to the data from Duncan et al. (1994) and the model fits to the data from the 3 subjects in Experiment 1. Figures 3 and 4 show the model fits to the data from the 3 subjects in Experiment 2. The estimated parameters for all fits are listed in Table 2, together with measures of goodness of fit (root mean squared deviations [RMSD]).

The model was fitted with encouraging precision to the data in Experiment 1, and the model also made good predictions of the data in Experiment 2 and Duncan et al. (1994). The largest RMSD values were found for the fits to the data in Experiment 2. This is not surprising, since the data in Experiment 2 were substantially more complex than the data in Experiment 1. In Experiment 1, only the SOA was varied, whereas in Experiment 2, both the SOA and the exposure duration of the targets were varied. On the other hand, the data in Experiment 2 constrained the model better than did the data in Experiment 1.

On the whole, the estimates for the parameters seem plausible. The estimated values of the total processing capacity, $C = 14.4 - 92.8$ Hz, are in the same range as previously estimated values ($C = 45$ Hz, Shibuya & Bundesen, 1988; $C = 23 - 26$ Hz, Finke et al., 2005; $C = 70$ Hz, Vangkilde, Bundesen, & Coull, 2011). The estimated values of the mean of t_0 , $\mu_0 = 8.7 - 29.5$ ms, are also consistent with previous findings ($t_0 = 18$ ms, Shibuya & Bundesen, 1988; $t_0 = 16 - 36$ ms, Finke et al., 2005; $t_0 = 15$ ms, Vangkilde et al., 2011). The remaining parameters in the model have not previously been explored, but the estimates seem realistic. For all subjects, the mean lock-time, μ_l , was estimated to be much shorter than the mean dwell-time, μ_d , so the subjects seemed much faster at locking neurons than at releasing them.

Another interesting observation is that the standard deviation of t_0 , σ_0 , is estimated to be larger for the subjects in Experiment 1, as compared with the subjects in Experiment 2 and the experiment by Duncan et al. (1994). This may relate to the masking of the letters. In Experiment 1, the masks were randomly generated, whereas in Experiment 2, 26 different masks were used. In Duncan et al., only one fixed mask was used. As touched upon earlier, the race toward VSTM is initiated t_1 ms after the onset of a letter

and lasts until the mask destroys the representation of the letter t_2 ms after the onset of the mask. Parameter t_0 is defined as the difference between t_1 and t_2 , $t_1 - t_2$. This means that t_0 is modulated by the effectiveness of the masks. That is, t_0 will be longer if the mask is effective, so that it quickly destroys the representation of the letter. On the other hand, if the mask is less effective, t_0 will decrease and may become negative. When only a single mask is used, the effectiveness of the mask will be the same on all trials, resulting in only a small variation in t_0 and a low estimate of σ_0 . However, when different masks are used, some will be more effective than others, and the variation in t_0 becomes larger and, thus, the estimate of σ_0 increases. In Experiment 2, the 26 different masks seemed nearly equally effective. By contrast, the random generation of masks in Experiment 1 resulted in a large variation in the effectiveness of the masks. This is probably the reason why the estimate of σ_0 is larger for the subjects in Experiment 1, as compared with the subjects in Experiment 2.

With as many as five free parameters, a natural question to ask is whether a model with fewer parameters might fit the data with the same precision. Two alternative models are interesting in this regard: A model in which $\sigma_0 = 0$ and a model in which $\mu_l = \mu_d$. Figure 5 (top left) shows the fits of these two alternative models to the data from subject 1 in Experiment 1. From visual inspection of the fits, the two models seem to perform worse than the TTVA model (with five free parameters). However, such a comparison must take the number of free parameters (i.e., the flexibility of the model) into account alongside the goodness of fit. For this reason, we employed the second-order Akaike and Bayesian information criteria (AICc, Sugiura, 1978, and Hurvich & Tsai, 1989; BIC, Schwarz, 1978), which penalize a model for additional free parameters. We computed AICc and BIC from least-squares statistics. That is,

$$\text{AICc} = n \ln \left(\frac{\sum \hat{\varepsilon}^2}{n} \right) + \frac{2nk}{n - k - 1} \quad (9)$$

and

$$\text{BIC} = n \ln \left(\frac{\sum \hat{\varepsilon}^2}{n} \right) + k \ln(n) \quad (10)$$

where $\sum \hat{\varepsilon}^2$ is the residual sums of squares, n is the sample size, and k is the number of free parameters. Table 3 shows AICc and BIC values for the three models fitted to the data from the 3 subjects in Experiment 1 and the 3 subjects in Experiment 2. In 5 out of the 6 subjects, AICc and BIC were lower for the TTVA model, as compared with the model in which $\sigma_0 = 0$ and the model in which $\mu_l = \mu_d$. This indicates that the TTVA model should be preferred over

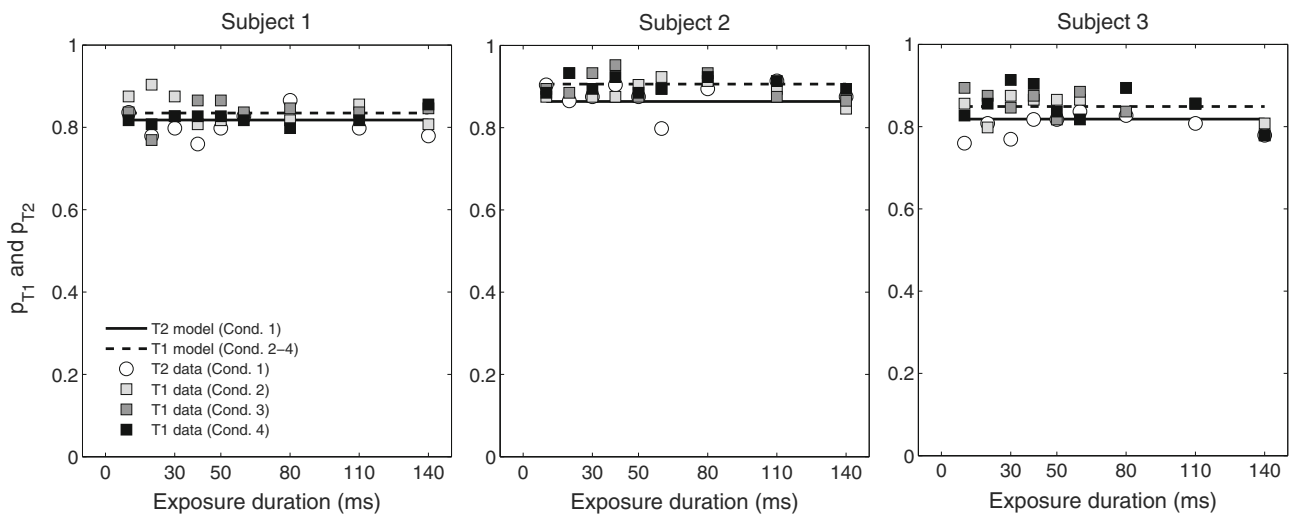


Fig. 3 Results of Experiment 2. The probability of correctly reporting T2 (p_{T2} , circles) as a function of exposure duration τ_1 of T1 in condition 1 (i.e., when τ_1 was varied), and the probabilities of correctly

reporting T1 (p_{T1} , squares) as functions of exposure duration τ_2 of T2 in conditions 2–4 (i.e., when τ_2 was varied). The flat lines are the model predictions

the two alternative models. F -tests comparing TTVA with the two nested models supported this conclusion. The F -tests revealed that TTVA fitted significantly better than the model in which $\sigma_0 = 0$, $F(6, 246) = 19.89$, $p < .001$, and the model in which $\mu_1 = \mu_d$, $F(6, 246) = 28.31$, $p < .001$.

In contrast, a competing model with the same number of parameters or even more parameters might exist. Here, an assumption of unlimited capacity is interesting in opposition to the capacity-limited assumption made in TTVA. TTVA can be made into an unlimited-capacity model by assuming that the pool of neurons is never exhausted but that the same number of neurons are available at all times. Figure 5a shows the best possible fit of such an unlimited capacity

model. Clearly, this model is not attractive, since the probability of correctly reporting T1 and T2 are the same at all SOAs. However, a mixture between a capacity-limited and a capacity-unlimited model might perform better. If p is the probability that on a given trial, the capacity is unlimited, the average processing rate will become $pC + (1 - p)v$, where C is the processing rate in the capacity-unlimited model and v is the processing rate in the capacity-limited model. We fitted this mixture model to the data from our two experiments and found that, on average, p had a value of .042 ($SD = .036$). That is, the mixture model provided almost the same fits as the capacity-limited model. Thus, TTVA with limited capacity should be preferred.

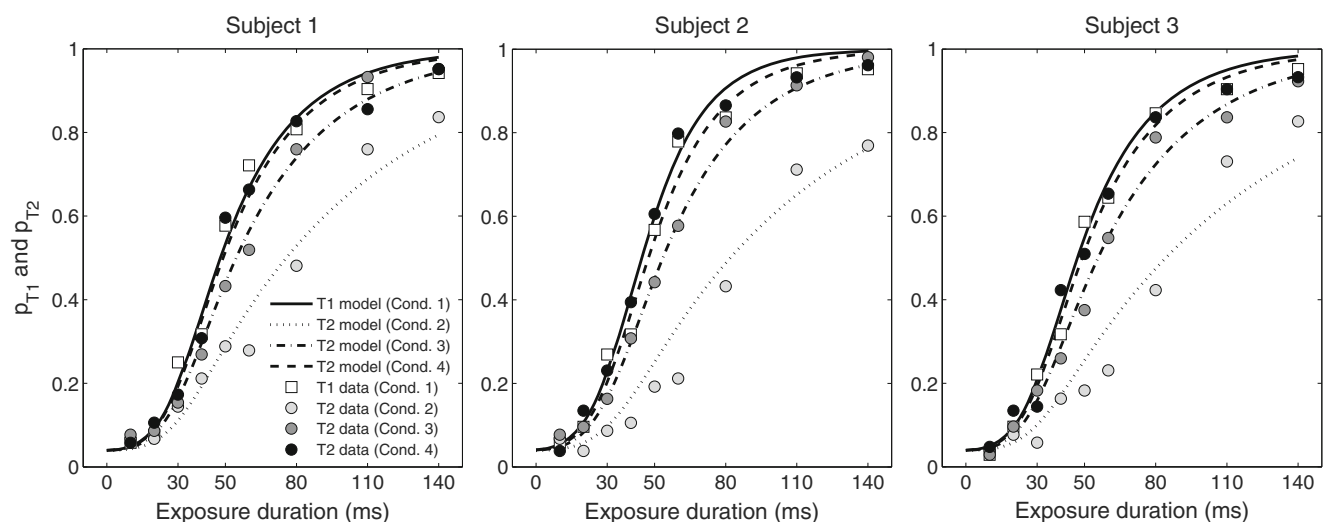


Fig. 4 Further results of Experiment 2: The probability of correctly reporting T1 (p_{T1} , squares) as a function of its exposure duration in condition 1, and the probabilities of correctly reporting T2 (p_{T2} ,

circles) as functions of its exposure duration in conditions 2–4. The sigmoid lines are the model predictions.

Table 1 Overview of the four different conditions in Experiment 2

Condition	τ_1 (ms)	τ_2 (ms)	SOA (ms)
1	10–140	80	900
2	80	10–140	200
3	80	10–140	500
4	80	10–140	900

Note. τ_1 and τ_2 indicate the exposure durations of T1 and T2, respectively, and SOA is the stimulus onset asynchrony between T1 and T2.

General discussion

We investigated the AD effect in two experiments in which we varied the SOA between two targets systematically between 0 and 900 ms (Duncan et al., 1994). In both experiments, we ran many trials across several experimental sessions to get reliable individual estimates of the probabilities of reporting either one of the two targets correctly. Trials with eye movements were discarded (cf. Petersen & Kyllingsbæk, 2012). In Experiment 1, we replicated and extended the findings of Duncan et al. (1994), keeping the exposure durations of the two targets (T1 and T2) constant at about 50 ms. Here, we found a fast decrease in the probability of correct report of T2 as SOA was increased from 0 to about 200 ms, followed by a gradual increase such that report of T2 was effectively equal to performance on T1 at the longest SOA of 900 ms. In Experiment 2, we ran a new version of the AD paradigm by varying not only SOA, but also the exposure durations of T1 and T2 systematically between 10 and 140 ms. We found strong effects of both exposure duration and SOA. T2 performance was again lowest at an SOA of 200 ms, and performance on T2 was similar to performance on T1 at the longest SOA of 900 ms.

A theory of temporal visual attention

We proposed a quantitative model to account for the results of the two experiments—a theory of temporal visual attention (TTVA). The model was based on the NTVA by Bundesen et al. (2005). As in NTVA, visual processing resources are distributed among objects such that the number of visual neurons representing an object is directly proportional to the attentional weight of the object. When the categorization "object x has feature i " enters VSTM, a node representing object x is activated in the topographic VSTM map of objects. To be retained in VSTM, some of the visual-processing neurons representing feature i of object x must become embedded in a positive feedback loop between these neurons and the node representing object x in the VSTM map. Thus, the perceptual machinery used to categorize stimuli in the visual field (neurons representing features of object x) is also utilized when information is retained in VSTM. The process of establishing the feedback

Table 2 Estimates of parameters by least-squares fits of TTVA to individual data from the 3 subjects in Experiment 1, group data reported by Duncan et al. (1994), and individual data from the 3 subjects in Experiment 2

	μ_l (ms)	μ_d (ms)	μ_0 (ms)	σ_0 (ms)	C (Hz)	RMSD
Exp 1, Sub 1	80.8	507.8	29.5	32.5	92.8	0.018
Exp 1, Sub 2	32.0	610.8	8.8	37.6	50.1	0.034
Exp 1, Sub 3	38.9	260.4	8.7	19.3	14.4	0.023
Duncan et al.	84.3	404.2	12.7	7.9	24.2	0.019
Exp 2, Sub 1	144.1	234.3	27.2	11.8	35.0	0.036
Exp 2, Sub 2	52.8	470.8	29.2	13.2	50.2	0.035
Exp 2, Sub 3	95.0	336.7	27.2	11.8	36.8	0.040

Note. Parameter μ_l is the mean lock-time, μ_d is the mean dwell-time, μ_0 is the mean of t_0 (the longest ineffective exposure duration), σ_0 is the standard deviation of t_0 , C is the total processing capacity, and RMSD is the square root of the mean squared deviation between observed and theoretical probabilities of correctly reporting T1 and T2.

loops of VSTM takes time. We model this time by a new rate parameter μ_l . When visual-processing neurons are locked for items encoded in VSTM, they cannot be used to process other stimuli in the visual field. This explains the attentional dwell time phenomenon. We conjecture that visual processing neurons are released from VSTM when the information has been recoded to a nonvisual (e.g., auditory, motoric, or amodal) format. The recoding process also takes time, and we modeled the rate of the recoding process by parameter μ_d . Thus, we modeled the AD phenomenon using two well-motivated parameters in addition to the parameters already given in NTVA. The model fitted the data of the two experiments with great precision.

Related work

Computational models of spatial shifts of visual attention Reeves and Sperling (1986; see also Sperling & Reeves, 1980) proposed an *attention gating model* (AGM) to account for the time course of spatial shifts of attention. Using the AGM, they modeled data from a paradigm using RSVP. Two visual streams containing letters and digits, respectively, were presented to the left and right of fixation. The rate of presentation was varied between 4.6 and 13.5 Hz. The task of the subject was to detect a predesignated target letter in the left stream and then shift attention to the right stream as quickly as possible to report the digits presented simultaneously with and following the target. In the AGM, an attentional gating function is modeled using a delayed gamma distribution comprising a convolution of two identical exponential distributions. Reeves and Sperling (1986) fitted the distribution to their data from 3 subjects and found rates of 7.52 Hz ($\mu = 133$ ms), 6.21 Hz ($\mu = 161$ ms), and 5.46 Hz

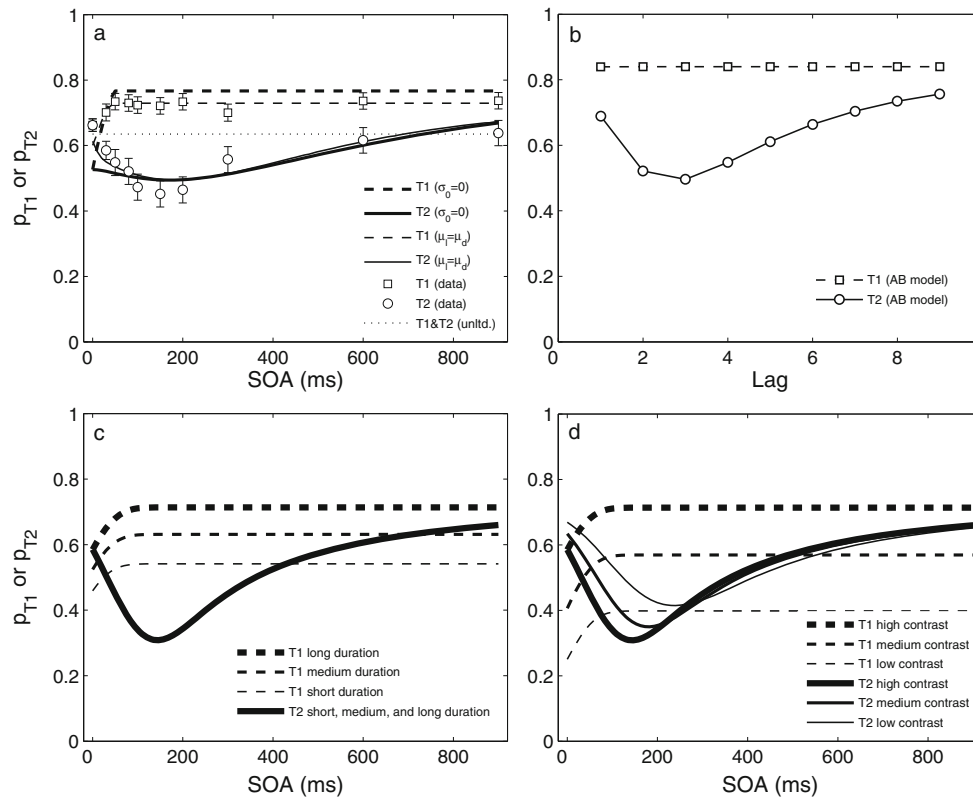


Fig. 5 Alternative models and model predictions. In all four panels, model predictions of p_{T1} and p_{T2} as functions of SOA are given by dashed and solid lines, respectively. **a** Fits of three alternative models (i.e., the model in which $\sigma_0 = 0$, thick lines; the model in which $\mu_1 = \mu_d$, thin lines; and the unlimited capacity model, dotted lines) to the data from subject 1 in Experiment 1. **b** Model predictions of p_{T1} and p_{T2} as a function of lag between the two targets in an AB paradigm with a rate of 100 ms per item. **c** Model predictions of p_{T1} and p_{T2} as functions of SOA when the exposure

duration of T1 is long (50 ms, thick line), medium (30 ms, medium thick line), or short (10 ms, thin line). The exposure duration of T2 was long (50 ms) in all three conditions. **d** Model predictions of p_{T1} and p_{T2} as functions of SOA when the contrast of T1 is high ($1.0 \times$ all η -values, thick line), medium ($0.5 \times$ all η -values, medium thick line), or low ($0.25 \times$ all η -values, thin line). The contrast of T2 was high ($1.0 \times$ all η -values) in all three conditions. In both panels c and d, estimated parameters from the fit to the data from subject 2 in Experiment 1 were used

($\mu = 183$ ms), respectively. Comparing these values with the estimates of μ_1 and μ_d derived from our model, our estimate of μ_1 was somewhat lower, whereas our estimate of μ_d was higher, but the estimates were similar in order of magnitude.

(1986) into an *episodic theory of the dynamics of spatial attention* (ETDSA), which describes the time course of visual attention as a sequence of discrete attentional episodes. The smooth transition between attentional episodes is described by a temporal transition function that is identical to the attentional (gamma) gating function of the AGM. In ETDSA, attention is analogous to a spotlight that illuminates only a single location

Sperling and Weichselgartner (1995; see also Weichselgartner & Sperling, 1987) extended the AGM of Reeves and Sperling

Table 3 AICc and BIC values for the three models (i.e., the TTVA model, the model in which $\sigma_0 = 0$, and the model in which $\mu_1 = \mu_d$) fitted to the data from the 3 subjects in Experiment 1 and the 3 subjects in Experiment 2

	AICc			BIC		
	TTVA	$\sigma_0 = 0$	$\mu_1 = \mu_d$	TTVA	$\sigma_0 = 0$	$\mu_1 = \mu_d$
Exp 1, Sub 1	-145.45	-104.70	-133.92	-144.75	-103.38	-132.60
Exp 1, Sub 2	-120.99	-115.40	-103.76	-120.29	-114.09	-102.45
Exp 1, Sub 3	-136.57	-136.26	-127.99	-135.87	-134.94	-126.68
Exp 2, Sub 1	-469.51	-455.99	-470.18	-459.03	-447.48	-461.67
Exp 2, Sub 2	-470.30	-442.86	-404.42	-459.83	-434.36	-395.91
Exp 2, Sub 3	-451.49	-440.22	-437.08	-441.02	-431.71	-428.57

Note. Highlighted numbers indicate the lowest AIC and BIC values for each subject.

at any given time (except in the period when attention is moved from one location to the next, decreasing at the old location and increasing at the new location). By contrast, TVA assumes that attention can be engaged simultaneously at several spatially separated locations.

Computational models of the attentional blink The AD effect bears resemblance to the AB effect, which has been studied extensively (e.g., Broadbent & Broadbent, 1987; Chun & Potter, 1995; Raymond et al., 1992). To investigate the AB, two targets (e.g., letters) are embedded in a RSVP stream of distractors (e.g., digits). Typically, a presentation rate of about 10 items per second is used (i.e., 100 ms per item). A strong impediment in report of T2 (i.e., the AB) is observed when T2 is presented about 200 ms after the presentation of T1. The time course of the AD and AB are very similar. Ward et al. (1997) noted this and presented experimental evidence in a skeletal version of the AB paradigm where only T1 and T2 were presented in the RSVP stream, each followed by a pattern mask similar to the one used in the AD paradigm. There is, however, an additional phenomenon that seems to appear more clearly in the AB paradigm than in the AD paradigm. This phenomenon is called the *lag 1 sparing effect* and occurs when T2 is presented immediately after T1, resulting in a preserved high performance on T2 as if the onset of the AB is delayed.

Figure 5b shows that TTVA can produce a standard AB with reasonable parameters (i.e., $\mu_1 = 100$ ms, $\mu_d = 500$ ms, $\mu_0 = 10$ ms, $\sigma_0 = 10$ ms, and $C = 20$ Hz). But at this stage, TTVA is not able to produce lag 1 sparing; at lag 1, the performance on T2 is forced to be lower than the performance on T1, except when the AB is very short. TTVA does, however, explain why a higher performance on T2 is found at lag 1 (i.e., at an SOA = 100 ms) in the AB, as compared with the similar performance in the AD paradigm: In the AB paradigm, T1 and T2 are presented at the same location. Consequently, all neurons that have not been locked to T1, when T2 is presented, will be exclusively available for the processing of T2. In contrast, these neurons will be distributed equally among the mask for T1 and T2 in the AD paradigm, because the mask for T1 and T2 are presented at different spatial locations in this paradigm.

Several other computational models of the AB have been proposed. For example, Shih (2008) presented an *attentional cascade model* of the AB (see also Shih & Sperling, 2002), which was based on cognitive theories of the AB (e.g., Chun & Potter, 1995; Giesbrecht & Di Lollo, 1998; Jolicoeur & Dell'Acqua, 1999; Shapiro, Raymond & Arnell 1994). The model of Shih is somewhat similar to TTVA by ascribing the AB to limitations in encoding/consolidation processes. However, in contrast to TTVA, the model does not make any predictions regarding the neural processes involved in the AB.

Furthermore, Bowman and Wyble (2007) presented a *simultaneous type, serial token (ST²) model* inspired by the two-stage theory of Chun and Potter (1995). In stage 1, a parallel visual processing of the stimuli is performed 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 RSVP stream. The binding happens in stage 2 by activation of a blaster (similar to the activation of a *transient attentional enhancement*; Nakayama & Mackeben, 1989). According to the ST² model, the AB occurs because the blaster is temporally suppressed until T1 is bound to a token and consolidated into VSTM, leaving T2 susceptible to decay and interruption by distractors. In TTVA, however, a stimulus is encoded in VSTM if and when any categorization of the stimulus is encoded in VSTM. That is, TTVA has only one stage of encoding to VSTM, and consequently, the capacity limitation is located in this stage when feedback loops from VSTM lock neurons to represent already encoded stimuli.

A transient attentional enhancement, or *boost*, is also a central mechanism in the *boost and bounce model* of Olivers and Meeter (2008). When a target is presented in the RSVP stream, a transient excitatory feedback will boost the encoding of the target and the following items in the stream into VSTM. The boost will be strongest for the item following immediately after the target. Thus, if T2 is presented at lag 1, it will be spared due to the boost initiated by T1. However, if the following item is a distractor, the boost will trigger a strong transient inhibitory feedback (the *bounce*), preventing the distractor and the following items in the stream to be encoded into VSTM. Thus, if T2 is presented at lag 2, an AB will be observed. In contrast to the attentional cascade model, the ST² model, and TTVA, the boost and bounce model assumes no central capacity limitations or bottlenecks to explain the AB. However, similar to TTVA, it gives essential explanatory powers to feedback loops from VSTM back to the mechanism responsible for the encoding of stimuli into VSTM.

Although the attentional cascade model and the ST² model are computational models and somewhat similar to TTVA by ascribing the AB to central capacity limitations, they are much more complex than TTVA. Presumably, the difference in the complexity of the theories reflects difference in the complexity of the phenomena they describe. To account for the AD data presented in this article, no special mechanism explaining lag 1 sparing had to be incorporated into TTVA. However, such a mechanism may have to be included if, at some point, TTVA is extended to explain lag 1 sparing as it occurs in the AB paradigm.

Predictions

A number of studies have examined how manipulation of T1 processing difficulty modulates the performance on T2. McLaughlin et al. (2001) manipulated T1 processing difficulty by reducing the exposure duration of T1 while keeping constant the duration of the target–mask complex. They found that this did not have any effect on the performance on T2, but only on how well T1 was reported. Figure 5c shows that this result is predicted by TTVA—the reason being that TTVA assumes that masks lock neurons in the same way as targets. Thus, reducing the exposure duration of T1 will make it more likely that the mask for T1 and not T1 will lock neurons; however, this does not affect the number of neurons available for processing of T2.

Chua (2005) used a different approach to manipulate T1 processing difficulty. By decreasing the luminance contrast of T1, Chua showed an attenuation of the impairment of correctly reporting T2. In TTVA, we may assume that all sensory evidence values (η -values) for features of an object decrease with the contrast of the object. Thus, it follows from the rate equation of TVA (see Eq. 1) that a decrease in contrast will result in a lower processing rate of the object. Moreover, the weight equation of TVA states that attentional weights are themselves derived from η -values (i.e., $w_x = \sum_{j \in R} \eta(x, j) \pi_j$ where R is the set of all visual categories, $\eta(x, j)$ is the strength of the sensory evidence that object x belongs to category j , and π_j is the pertinence of category j). Consequently, the attentional weight of an object will decrease with the contrast of the object. Figure 5d shows the predictions made by TTVA when contrast is varied. In line with Chua, TTVA predicts that lowering the contrast of T1 results in an attenuation of the impairment in correctly reporting T2.

Finally, as was mentioned in the introduction, Moore et al. (1996) found that removing the mask for T1 reduced the duration of the AD effect. TTVA predicts this finding: As has previously been mentioned, TTVA assumes that masks compete for and lock neurons in the same way as targets.

Thus, removing the mask for T1 will leave T2 without competition after the offset and decay of T1 resulting in faster recovery of T2.

Conclusion

We have proposed a quantitative model accounting for the AD phenomenon based on highly accurate measures of its time course. The core assumption in the model is that retention of a stimulus in VSTM takes up visual-processing resources used to encode the stimulus into VSTM. Thus, retention of the first target in the AD paradigm leads to a temporary lack of available processing resources, which explains the observed impairment in correctly reporting the second target.

Author Note Anders Petersen, Søren Kyllingsbæk, and Claus Bundesen, Center for Visual Cognition, Department of Psychology, University of Copenhagen, Copenhagen, Denmark.

We thank Simon Nielsen for collecting and analyzing part of the data in this article.

Correspondence concerning this article should be addressed to Anders Petersen, Center for Visual Cognition, Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen K, Denmark. E-mail: anders.petersen@psy.ku.dk.

The research was supported by the Danish Council for Independent Research, the Danish Council for Strategic Research, and the University of Copenhagen.

Appendix

When two letters, T1 and T2, are presented with a temporal gap (SOA), we assume that two redistributions of the attentional resources (neurons) will occur—one at t_{01} (i.e., t_0 for T1) and one at t_{02} (i.e., t_0 for T2). Furthermore, we assume that t_{01} and t_{02} are approximately normally distributed with means μ_0 and SOA + μ_0 , respectively, and standard deviation σ_0 . If $t_{01} \leq t_{02}$, the processing rate of T2 is given by

$$\begin{aligned}
 v_{T2}(t_{01} \leq t_{02}) &= C \left(\frac{w_{T2}}{w_{T1} + w_{T2}} \times p_{\text{free}}|(t_{01} \leq t_{02}) + 0 \times p_{\text{locked}}|(t_{01} \leq t_{02}) + 1 \times p_{\text{released}}|(t_{01} \leq t_{02}) \right) \\
 &= C \left(\frac{1}{2} p_{\text{free}}|(t_{01} \leq t_{02}) + p_{\text{released}}|(t_{01} \leq t_{02}) \right), \tag{11}
 \end{aligned}$$

where we have made the plausible assumption that $w_{T1} = w_{T2}$. Here, C is the total processing capacity, p_{free} is the proportion of neurons that become distributed according to the attentional weights, p_{locked} is the proportion of neurons that remain locked to T1, and p_{released} is the proportion of neurons that have been released from T1 and are exclusively available for T2.

However, p_{free} , p_{locked} , and p_{released} also represent the probabilities with which a single neuron is found in the three stages. Thus, the proportions above can be derived by defining the behavior of a single neuron. The times it takes to lock and release a single neuron are assumed to be exponentially distributed with rate parameters

λ_1 and λ_d , respectively. If we furthermore assume that the locking process starts immediately after an object has been encoded into VSTM and the subsequent re-

lease process starts directly after a neuron has been locked in a feedback loop, the three proportions (probabilities) are given by

$$p_{\text{free}}|(t_{01} \leq t_{02}) = 1 - \int_{t_{01}}^{t_{02}} C e^{-C(t-t_{01})} \left(\int_t^{t_{02}} \lambda_1 e^{-\lambda_1(t'-t)} dt' \right) dt = e^{C(t_{01}-t_{02})} + \frac{C}{\lambda_1 - C} \left(e^{C(t_{01}-t_{02})} - e^{\lambda_1(t_{01}-t_{02})} \right) \quad (12)$$

$$\begin{aligned} p_{\text{locked}}|(t_{01} \leq t_{02}) &= \int_{t_{01}}^{t_{02}} C e^{-C(t-t_{01})} \left(\int_t^{t_{02}} \lambda_1 e^{-\lambda_1(t'-t)} \left(1 - \int_{t'}^{t_{02}} \lambda_d e^{-\lambda_d(t''-t')} dt'' \right) dt' \right) dt \\ &= \frac{\lambda_1}{\lambda_d - \lambda_1} \left(\frac{C}{\lambda_1 - C} \left(e^{C(t_{01}-t_{02})} - e^{\lambda_1(t_{01}-t_{02})} \right) - \frac{C}{\lambda_d - C} \left(e^{C(t_{01}-t_{02})} - e^{\lambda_d(t_{01}-t_{02})} \right) \right) \end{aligned} \quad (13)$$

$$\begin{aligned} p_{\text{released}}|(t_{01} \leq t_{02}) &= \int_{t_{01}}^{t_{02}} C e^{-C(t-t_{01})} \left(\int_t^{t_{02}} \lambda_1 e^{-\lambda_1(t'-t)} \left(\int_{t'}^{t_{02}} \lambda_d e^{-\lambda_d(t''-t')} dt'' \right) dt' \right) dt \\ &= 1 - e^{C(t_{01}-t_{02})} - \frac{C}{\lambda_1 - C} \left(e^{C(t_{01}-t_{02})} - e^{\lambda_1(t_{01}-t_{02})} \right) \\ &\quad - \frac{\lambda_1}{\lambda_d - \lambda_1} \left(\frac{C}{\lambda_1 - C} \left(e^{C(t_{01}-t_{02})} - e^{\lambda_1(t_{01}-t_{02})} \right) - \frac{C}{\lambda_d - C} \left(e^{C(t_{01}-t_{02})} - e^{\lambda_d(t_{01}-t_{02})} \right) \right) \end{aligned} \quad (14)$$

where t is the time when the letter identity of T1 (or the mask for T1) is encoded into VSTM, t' is the time when a neuron representing the encoded identity of T1 is locked by becoming embedded in a positive feedback loop gated by the unit representing T1 in the VSTM map, and t'' is the time when the feedback loop is broken such that the neuron is released. At any given time, a neuron is in just one of the three states, so $p_{\text{free}} + p_{\text{locked}} + p_{\text{released}} = 1$. Figure 6 shows an example of how the above probabilities

change as functions of SOA in the simple case in which the time from the letter is presented until visual-processing resources are allocated to the letter is the same for T1 and T2 (i.e., $t_{02} = \text{SOA} + t_{01}$).

The above applies only when $t_{01} \leq t_{02}$. If SOA is short, the distribution of t_{02} overlaps the distribution of t_{01} , and it becomes likely that $t_{01} > t_{02}$. In this case, T2 will have a head start and be assigned a processing rate of C in the interval between t_{02} and t_{01} . After t_{01} , the processing rate of T2 will be

$$\begin{aligned} v_{T2}|(t_{01} > t_{02}) &= C \left(\frac{w_{T2}}{w_{T1} + w_{T2}} \times p_{\text{free}}|(t_{01} > t_{02}) + 1 \times p_{\text{locked}}|(t_{01} > t_{02}) + 0 \times p_{\text{released}}|(t_{01} > t_{02}) \right) \\ &= C \left(\frac{1}{2} p_{\text{free}}|(t_{01} > t_{02}) + p_{\text{locked}}|(t_{01} > t_{02}) \right), \end{aligned} \quad (15)$$

where p_{locked} is now the proportion of neurons locked to T2 and p_{released} is the proportion of neurons released from T2. As previously stated, p_{free} is the proportion of neurons distributed according to the attentional weights. These three proportions are calculated in the same way as when $t_{01} \leq t_{02}$, with the only difference that t_{01} and t_{02} are interchanged.

Given the calculated processing rates of T2 when $t_{01} \leq t_{02}$ [i.e., $v_{T2}|(t_{01} \leq t_{02})$] and $t_{01} > t_{02}$ [i.e., $v_{T2}|$

$(t_{01} > t_{02})$], the probability of encoding T2 (p_{T2}) is found as a sum of three probabilities: $p[\text{T2} \& (t_{01} \leq t_{02})]$, $p[\text{T2} \& (t_{01} > t_{02}) \& (t_{01} \leq \text{SOA} + \tau_2)]$, and $p[\text{T2} \& (t_{01} > t_{02}) \& (t_{01} > \text{SOA} + \tau_2)]$. In the first two conditions, the presentation of T1 affects the processing rate of T2. However, in the last condition, the presentation of T1 does not affect the processing rate of T2, resulting in a rate of C from t_{02} until $\text{SOA} + \tau_2$. Thus, the three probabilities are given by

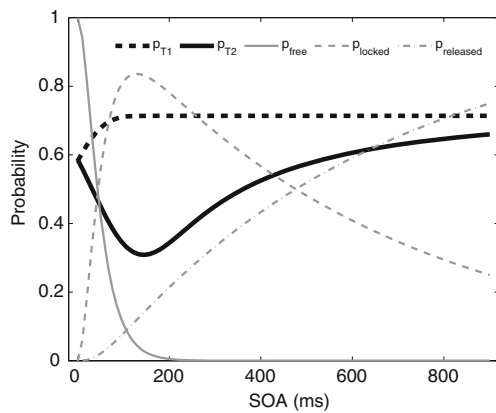


Fig. 6 Locking and releasing of resources. This example shows the model fit to the data from subject 2 in Experiment 1. The thick black lines show the probabilities of encoding T1 (p_{T1}) and T2 (p_{T2}) as functions of SOA and match the probabilities in Fig. 2. The thin gray lines show the underlying probabilities with which a visual processing neuron is free (p_{free}), locked to T1 (p_{locked}), or released from T1 ($p_{released}$) in the simple case in which the time from the letter is the same for T1 and T2 (i.e., $t_{02} = SOA + t_{01}$). Note that $p_{free} + p_{locked} + p_{released} = 1$

$$p[T2 \& (t_{01} \leq t_{02})] = \int_{-\infty}^{SOA+\tau_2} \frac{1}{\sigma_0} \phi\left(\frac{t_{01} - \mu_0}{\sigma_0}\right) \left(\int_{t_{01}}^{SOA+\tau_2} \frac{1}{\sigma_0} \phi\left(\frac{t_{02} - (SOA + \mu_0)}{\sigma_0}\right) \left(1 - e^{-[v_{T2}](t_{01} \leq t_{02})(SOA+\tau_2-t_{02})} \right) dt_{01} \right) dt_{01} \tag{16}$$

$$p[T2 \& (t_{01} > t_{02}) \& (t_{01} \leq SOA + \tau_2)] = \int_{-\infty}^{SOA+\tau_2} \frac{1}{\sigma_0} \phi\left(\frac{t_{02} - (SOA + \mu_0)}{\sigma_0}\right) \left(\int_{t_{02}}^{SOA+\tau_2} \frac{1}{\sigma_0} \phi\left(\frac{t_{01} - \mu_0}{\sigma_0}\right) \left(1 - e^{-C(t_{01}-t_{02}) - [v_{T2}](t_{01} > t_{02})(SOA+\tau_2-t_{01})} \right) dt_{01} \right) dt_{02} \tag{17}$$

$$p[T2 \& (t_{01} > t_{02}) \& (t_{01} > SOA + \tau_2)] = \left(1 - \Phi\left(\frac{SOA + \tau_2 - \mu_0}{\sigma_0}\right)\right) \int_{-\infty}^{SOA+\tau_2} \frac{1}{\sigma_0} \phi\left(\frac{t_{02} - (SOA + \mu_0)}{\sigma_0}\right) \left(1 - e^{-C(SOA+\tau_2-t_{02})}\right) dt_{02}, \tag{18}$$

and the probability of encoding T2 is given by

$$p_{T2} = p[T2 \& (t_{01} \leq t_{02})] + p[T2 \& (t_{01} > t_{02}) \& (t_{01} \leq SOA + \tau_2)] + p[T2 \& (t_{01} > t_{02}) \& (t_{01} > SOA + \tau_2)]. \tag{19}$$

The probability of encoding T1 (p_{T1}) is found by a similar calculation and given by the following sum:

$$p_{T1} = p[T1 \& (t_{02} \leq t_{01})] + p[T1 \& (t_{02} > t_{01}) \& (t_{02} \leq \tau_1)] + p[T1 \& (t_{02} > t_{01}) \& (t_{02} > \tau_1)]. \tag{20}$$

References

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review*, *114*, 38–70.

Brehaut, J. C., Enns, J. T., & Di Lollo, V. (1999). Visual masking plays two roles in the attentional blink. *Perception & Psychophysics*, *61*, 1436–1448.

Broadbent, D. E., & Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*, 105–113.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, *112*, 291–328.

Bundesen, C., Pedersen, L. F., & Larsen, A. (1984). Measuring efficiency of selection from briefly exposed visual displays: A model for partial report. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 329–339.

Bundesen, C., Shibuya, H., & Larsen, A. (1985). Visual selection from multielement displays: A model for partial report. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 631–649). Hillsdale, NJ: Erlbaum.

Chua, F. K. (2005). The effect of target contrast on the attentional blink. *Perception & Psychophysics*, *67*, 770–788.

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*, 109–127.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517.

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

Dyrholm, M., Kyllingsbæk, S., Espeseth, T., & Bundesen, C. (2011). Generalizing parametric models by introducing trial-by-trial parameter variability: The case of TVA. *Journal of Mathematical Psychology*, *55*, 416–429.

Finke, K., Bublak, P., Krummenacher, J., Kyllingsbæk, S., Muller, H. J., & Schneider, W. X. (2005). Usability of a theory of visual attention (TVA) for parameter-based measurement of attention: I. Evidence from normal subjects. *Journal of the International Neuropsychological Society*, *11*, 832–842.

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

Hebb, D. (1949). *Organization of behavior*. New York: Wiley.

Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307.

Jolicoeur, P., & Dell’Acqua, R. (1999). Attentional and structural constraints on visual encoding. *Psychological Research*, *62*, 154–164.

- Jolicoeur, P., & Dell'Acqua, R. (2000). Selective influence of second target exposure duration and task-1 load effects in the attentional blink phenomenon. *Psychonomic Bulletin & Review*, 7, 472–479.
- Kesten, H. (1958). Accelerated stochastic approximation. *Annals of Mathematical Statistics*, 29, 41–59.
- Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, 4, 138–147.
- Kyllingsbæk, S., Schneider, W. X., & Bundesen, C. (2001). Automatic attraction of attention to former targets in visual displays of letters. *Perception & Psychophysics*, 63, 85–98.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- McLaughlin, E. N., Shore, D. I., & Klein, R. M. (2001). The attentional blink is immune to masking-induced data limits. *Quarterly Journal of Experimental Psychology*, 54A, 169–196.
- Moore, C. M., Egeth, H., Berglan, L. R., & Luck, S. J. (1996). Are attentional dwell times inconsistent with serial visual search? *Psychonomic Bulletin & Review*, 3, 360–365.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229, 782–784.
- Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 315–330.
- Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, 29, 1631–1647.
- Olivers, C. N. L., & Meeter, M. (2008). A boost and bounce theory of temporal attention. *Psychological Review*, 115, 836–863.
- Petersen, A., & Kyllingsbæk, S. (in press). Eye movements and practice effects in the attentional dwell time paradigm. *Experimental Psychology*.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.
- Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (pp. 531–556). Hillsdale, NJ: Erlbaum.
- Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, 81, 10–15.
- 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.
- Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Research*, 93, 180–206.
- Schneider, W., & Fisk, A. D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. *Perception & Psychophysics*, 31, 160–168.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Shapiro, K., 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, 357–371.
- Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 591–600.
- Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology*, 93, 72–82.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127–190.
- Shih, S.-I. (2008). The attention cascade model and attentional blink. *Cognitive Psychology*, 56, 210–236.
- Shih, S.-I., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review*, 109, 260–305.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74, 1–29.
- Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica*, 27, 285–292.
- Sperling, G., & Reeves, A. (1980). Measuring the reaction time of a shift of visual attention. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 347–360). Hillsdale, NJ: Erlbaum.
- Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review*, 102, 503–532.
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics-Theory and Methods*, 7, 13–26.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Vangkilde, S., Bundesen, C., & Coull, J. T. (2011). Prompt but inefficient: Nicotine differentially modulates discrete components of attention. *Psychopharmacology*, 218, 667–680.
- Ward, R., Duncan, J., & Shapiro, K. (1996). The slow time-course of visual attention. *Cognitive Psychology*, 30, 79–109.
- Ward, R., Duncan, J., & Shapiro, K. (1997). Effects of similarity, difficulty, and nontarget presentation on the time course of visual attention. *Perception & Psychophysics*, 59, 593–600.
- Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science*, 238, 778–780.
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238.