MIT Open Access Articles

Attentional episodes in visual perception

Massachusetts Institute of Technology

Attentional episodes in visual perception

Brad Wyble[1], **Mary C Potter**[2], **Howard Bowman**[3], and **Mark Nieuwenstein**[4]

[1]Syracuse University, Syracuse NY [2]MIT, Cambridge, MA [3]University of Kent, Canterbury, UK [4]University of Groningen, Groningen, the Netherlands

Abstract

Is one's temporal perception of the world truly as seamless as it appears? This paper presents a computationally motivated theory suggesting that visual attention samples information from temporal episodes (episodic Simultaneous Type/ Serial Token model or eSTST; Wyble et al 2009a). Breaks between these episodes are punctuated by periods of suppressed attention, better known as the attentional blink (Raymond, Shapiro & Arnell 1992). We test predictions from this model and demonstrate that subjects are able to report more letters from a sequence of four targets presented in a dense temporal cluster, than from a sequence of four targets that are interleaved with non-targets. However, this superior report accuracy comes at a cost in impaired temporal order perception. Further experiments explore the dynamics of multiple episodes, and the boundary conditions that trigger episodic breaks. Finally, we contrast the importance of attentional control, limited resources and memory capacity constructs in the model.

Keywords

attentional blink; attention; episode; lag-1 sparing; working memory

Introduction

Our waking perception of the world is apparently continuous much of the time, yet the psychology of temporal attention suggests that our attentional state fluctuates at rapid time scales. Beginning with the theory of attention waves that peak at 1.5 second intervals, as described by Titchener (1910, see also Pechenkova 2006), the idea of temporal fluctuations in our readiness to perceive visual input has been critical to our understanding of perception and awareness. In this paper we describe a theory relating the fine grained time course of these fluctuations to the temporal structure of incoming visual stimuli.

A useful experimental paradigm for studying temporal attention involves rapid serial visual presentation experiments (RSVP), in which stimuli are presented rapidly, in sequence, and with each stimulus replacing the previous one. RSVP allows experiments to focus on the temporal dynamics of processing target items (i.e. stimuli that subjects are attempting to perceive and report) presented among a sequence of irrelevant distractor items. A major finding of interest in such experiments is that there are rapid transitions in the attentional

state of a participant in response to target items.. This attentional blink reveals rapid fluctuations in the ability to report target stimuli within short intervals similar to the durations of eye fixations (Broadbent & Broadbent 1987; Weichselgartner & Sperling 1986; Raymond, Shapiro & Arnell 1992; c.f. Chun & Potter 1995; Martens & Wyble 2010). In such paradigms, the typical attentional blink finding is one of impaired report of a second target (T2) when it appears 200–500 ms after the onset of the first target (T1) .

With respect to this attentional blink, it is somewhat paradoxical that if two targets are presented even more closely together (e.g. onsets within about 100 ms), subjects frequently report both of them, thereby demonstrating an effect known as lag 1 sparing (Potter, Chun, Banks, & Muckenhoupt, 1998; see Visser et al., 1999, for a review). This sparing effect has more recently been found for a string of 3 or 4 targets (Di Lollo, Kawahara, Ghorashi & Enns 2005; Olivers, van der Stigchel & Hulleman 2007; Kawahara, Kumada & Di Lollo 2006), or across a sequence of 5 items or more in unselective processing (Nieuwenstein & Potter 2006; Potter, Nieuwenstein & Strohminger, 2008). Such sparing effects have been enormously influential in defining the theoretical landscape of the attentional blink literature (Martens & Wyble 2010) and have provoked vigorous debate about the underlying mechanisms of attention (Dell Acqua, et al. 2009; Dux, Asplund & Marois, 2009; Olivers, Spalek, Kawahara, & Di Lollo, 2009; Olivers, Hulleman, Spalek, Kawahara & Di Lollo, In Press). We will address this debate in the discussion.

Beyond the attentional blink

A tremendous amount of emprical knowledge has been acquired about the temporal dynamics of human performance in attentional blink experiments. The acquisition of these data has coincided with the development of computational models of temporal attention which are directly inspired by the attentional blink and are becoming increasingly sophisticated with each iteration (e.g. Nieuwenhuis Gilzenrat, Holmes & Cohen 2005, Bowman & Wyble 2007, Wyble, Bowman & Nieuwenstein 2009a; Shih 2008, Olivers & Meeter 2009; Taatgen, Juvina, Schipper, Borst, & Martens 2009).

It is not the purpose of this paper to present the evidence for and against these competing accounts (c.f. Martens & Wyble 2010). Rather, in the present paper we expand the scope of one model, the episodic Simultaneous Type/ Serial Token (eSTST) model (Wyble, et al. 2009a), beyond the bounds of the attentional blink phenomenon and into a broader theory about the time course of visual attention during a dynamic stream of visual input, such as an RSVP stream containing a sequence of four or more targets. In particular, the model suggests that the visual system is predisposed to allocating attention in uninterrupted temporal packets with endpoints defined by the timing of task relevant stimuli. Visual targets presented in close temporal proximity are readily encoded, while targets that are spread farther apart in time will be encoded less often.

The eSTST model simulates an attentional mechanism that is inherently episodic by which we mean that the model is best suited to processing visual input that arrives in temporal chunks that are at least 200 milliseconds in length and can be extended in response to the arrival of additional targets. This paper starts with the computational model exactly as formulated by Wyble et al. (2009a) and describes a series of predictions for a set of novel experiments that define the construct of episodic attention. These experiments will elucidate the benefits (prediction 1) and costs (prediction 3) of episodic encoding, the boundary conditions of terminating such episodes (predictions 2,6), the ability to encode multiple episodes (prediction 4), the consequences of encoding items within or between episodes (prediction 5) and the role of overall capacity limits (prediction 7) in determining performance in these tasks. Testing novel predictions of an extant model, as we will do here, is the most effective way to assess the validity of a computational theory. If the predictions

are matched by the data, we infer that there is truth value in the simulated mechanisms of the model. A further strength of this approach is that neurophysiologically inspired models, such as eSTST, provide a mechanistic explanation of the cognitive processes that they simulate.

The eSTST model was originally based on a simulation of several findings in RSVP, including the attentional blink, whole report (Nieuwenstein & Potter 2006) and repetition blindness (Kanwisher 1987, Mozer 1989). The heart of the theory embodied by the model is that the cognitive circuitry that produces attentional blinks is actually designed to help us parse visual input into temporal packets as they are encoded into memory. To simulate this process, we propose a competitive regulatory circuit which is *highly sensitive to the endpoint of a sequence of salient stimuli*. At such a breakpoint, attention is suppressed to hold off new information until the just-acquired stimuli are successfully encoded into working memory (Figure 1). This circuit produces temporal fluctuations in the intake of new information which we refer to as *attentional episodes*.i

What is an attentional episode?

We define an attentional episode as a temporal interval during which attention remains strongly engaged, allowing one or more targets to enter the encoding process. In the context of search through a serial sequence of stimuli, an episode is triggered by an initial target and is sustained until there is a temporal gap with no new targets that is sufficiently long to terminate the deployment of attention. In the case of RSVP, the critical duration of this gap seems to be about 150msii. For example, in an RSVP stream with items presented for 100 ms, a single target in isolation is encoded as a single episode. If two targets are separated by a single distractor (i.e., a 200 ms Target Onset Asynchrony or TOA), the attentional episode is terminated after the first target, and attention is suppressed until encoding of the items acquired during the first episode is complete, thus producing an attentional blink for the second target. If two or more targets are presented in succession at 100ms TOA, the episode is extended to contain the entire sequence, which might include three or more RSVP targets (Di Lollo, Kawahara, Ghorashi & Enns 2005; Olivers van der Stigchel & Hulleman 2007; Kawahara, Kumada & Di Lollo 2006; Wyble et al 2009a).

There are two crucial assumptions in this theoretical framework. The first is that encoding of an item into memory, and perhaps even into visual awareness, can substantially outlast the physical duration of that stimulus, even when that stimulus has been backward masked. The second is that this encoding process can operate on multiple items simultaneously, although it may sometimes be advantageous to force two items to be encoded sequentially, rather than simultaneously. The mediation between these two methods of encoding, *simultaneous and sequential*, results in the behavioral phenomena known as the attentional blink, and lag-1 sparing.

The idea of delaying attention to defer encoding of new stimuli is similar to the delayed attentional engagement theory described by Nieuwenstein, Chun, van der Lubbe, & Hooge, (2005) and is supported by evidence of delayed processing of a T2 during the AB (Vogel &

---

iSperling & Weichselgartner (1995) proposed an episodic theory of attentional gating. In their terminology, episodes were the quantal, discrete shifts of attention from one mode of processing (i.e. detecting targets -> encoding targets -> closed attentional gate) or from one spatial location to another. Such work by Sperling and colleagues (including also Reeves & Sperling 1986) has motivated the understanding of temporal changes in attentional state that we describe here. The terminological differences can be reconciled by understanding that our definition of an episode refers only to periods of time when the attentional gate is open and encoding is therefore facilitated. Our definition of an episode is very similar to Sperling & Weichselgartner's (1995) description of an open attentional gate.
iiThe estimate of a critical gap length of about 150 ms is derived from letter/digit RSVP experiments and may vary for different stimuli, although initial experiments with picture stimuli in RSVP tasks (Evans & Treisman 2005; Potter, Wyble, Pandav, & Olejarczyk 2010) suggest that picture processing has similar temporal characteristics with regard to the attentional blink and lag 1 sparing.

Luck 2002; Vul, Nieuwenstein & Kanwisher 2008, Bowman, Wyble, Chennu & Craston 2008). This idea also shares ground with the notion of wrap-up time at a clause boundary (Haberlandt & Graesser 1989; Rayner, Kambe & Duffy 2000) or the end of a sentence (Just & Carpenter 1980) in reading behavior.

The role of episodes in perception

We hypothesize that the purpose of an attentional mechanism with episodic behavior is to bundle together temporally proximal stimuli into mnemonic chunks, while at the same time ensuring that stimuli separated by temporal gaps are not grouped together. The underlying rationale for this explanation is that there exists a fundamental tradeoff between our ability to rapidly encode stimuli, and their episodic distinctiveness (Wyble et al 2009a). For example, when viewers report both targets during lag 1 sparing they exhibit an exaggerated proportion of temporal swap and misbinding errors (Akyürek, & Hommel 2005; Hommel & Akyurek 2005; Bowman & Wyble 2007; Chun & Potter 1995; Wyble et al 2009). In the context of attentional episodes, these temporal order and binding errors are the result of encoding items together within the same episode. Likewise, repetition blindness (the failure to see the second of two repeated items presented at a short lag: Kanwisher 1987, Mozer 1989) reflects the difficulty of encoding two items within a single episode when both are identical. For example, when subjects see a sequence of 4 targets in which the first and the fourth are identical (i.e. the sequence is TTTR, where T is a target and R is a repetition of the first target) severe repetition blindness is obtained for the fourth target (Wyble et al. 2009a). However, separating the two identical targets with distractors (i.e. TDDR where D represents a distractor item) dramatically reduces the incidence of repetition blindness. In simulations of attentional episodes (Wyble et al. 2009a), the TDDR sequence allows the initial presentation of the target and its repetition to be encoded in separate episodes.

According to this theory of episodic attention, visual input presented with temporally clustered targets should be better perceived than information that is more separated in time. To be concrete, consider the task of selecting *letter targets* in a background of *digits* in RSVP. The episodic theory predicts that more items will be reported from a sequence of letters if they are presented in a single cluster, such as 853HKMG95, or separated into two clusters such as 8HK53968MG23, rather than interleaved with distractors, such as 8H5K3M9G5. However, the greater number of reported items in the clustered condition will coincide with a reduced ability to correctly report their sequential order and increased repetition blindness.

The eSTST Model

In this paper we test this theory of attentional episodes. We begin by describing the model originally presented in Wyble et al. (2009a) and then report new experimental work which evaluates the aforementioned predictions of the model. Code for running the simulations can be found at http://www.bradwyble.com/research/models/eSTST

eSTST: A Model of Episodic Attention

The episodic Simultaneous Type/ Serial Token model (eSTST; Wyble, et al. 2009a) and its predecessor the Simultaneous Type/Serial Token model (Bowman & Wyble 2007) describe the interaction between attention and working memory. These are both neural networks built on principles of basic neurophysiology, and they simulate the deployment of temporal attention as well as the encoding of information into working memory. In this paper, we use the eSTST model described in the original publication without modification of parameters except for one case in Experiment 3, and the addition of two additional tokens to accommodate 6-target sequences.

Competitive Regulation of Attention—The eSTST model simulates the temporal dynamics of visual attention with a circuit that resolves competition between top down and bottom up influences on a central attentional node (Figure 2). This node acts as a gate that can help visual stimuli enter working memory by amplifying the strength (i.e., adjusting the gain; Reeves & Sperling 1986) of forward going connections leading from early visual representations to later stages of processing. The attentional gate receives both *excitation* from salient stimuli (e.g., targets in the current task, one's own name, emotionally valent stimuli), and *inhibitory control* from central processes that are engaged once a stimulus has entered the encoding stage. Keeping the metaphor of the attentional gate (Reeves & Sperling 1986), once attention has been triggered, it is nonselective, enhancing the processing of all stimuli equally. This model is designed to simulate central RSVP tasks with no spatial component; however, it is assumed that the attentional gate is spatially specific (Wyble, Bowman & Potter 2009).

Continuing attentional enhancement—The attention control circuit has the property of excitatory recurrence: targets excite attention, and attention in turn amplifies the processing of targets. Consequently, a string of briefly presented targets provides an especially potent input to attention, with each target giving an attentional boost to the next target.

Inhibitory control—A target that has entered the encoding stage generates a working memory representation over the course of several hundred milliseconds. During this period of encoding the mechanisms responsible for memory encoding *attempt* to suppress attention. If no additional targets follow, suppression succeeds in shutting down attention, thus causing an attentional blink that persists until T1 encoding is complete (Figure 3a). However, if an additional target immediately follows T1, that target is amplified by attention and can hold the attentional gate open (Figure 3b) despite the inhibitory control. Therefore, the net effect of this inhibitory control is *to delay the encoding of new information following a gap in the target sequence*. This delay increases the chance that stimuli in the first episode can be completely encoded before attention can be reactivated.

A Type/Token Account of Working Memory

To understand the benefit of encoding stimuli into distinct episodes, it is helpful to consider types and tokens as a memory substrate. The STST model (Bowman & Wyble 2007) and the more recent eSTST model (Wyble, et al. 2009a) simulate the encoding of items into working memory by creating a binding link between a representation of the item's identity (its *type)* and a representation of the spatiotemporal event, called a *token*. In the model, the binding link represents the identity of the stimulus and the token represents the fact that a given type appeared at a particular time and place. The tokens are stored in a serial, temporal sequence, thus allowing the model to recover both the identity and the temporal order of the presented targets[iii]. An additional characteristic of a tokenized working memory is the ability to store repetitions--different instances of the same type. See Mozer (1989), Kanwisher (1989), Chun (1997) and Bowman and Wyble (2007) for further discussion of the *type/token* framework.

The complete eSTST model is depicted in Figure 4. Each circle in the diagram corresponds to a node in the neural network, and these nodes are connected by fixed excitatory or inhibitory weights that are not modifiable. All units are threshold linear, which means that their output is a linear function of the degree to which each unit's activity exceeds its

---

[iii]This definition of a token differs somewhat from that described in the STST model by Bowman & Wyble 2007, in which lag-1 sparing was thought to represent 2 items being bound to a single token. In the eSTST model, a token always represents 1 item.

threshold, which ranges from 0 to a positive number. See the original Wyble et al (2009a) for a description of the full set of equations and parameters.

Processing in the model depicted in Figure 4 proceeds generally from bottom to top. At the bottom of the figure, input nodes extract type information (i.e., an abstract identity representation of each target) and simulate backward masking with feedforward inhibition. When a visual stimulus is presented to the model, it activates a corresponding input node. These input nodes are filtered by a passive task demand which is configured by the task requirements to suppress distractors and allow targets to pass into the type layer of the model. Thus, in the eSTST model, the filtering of targets from distractors is accomplished by a passive mechanism that does not interact with the temporal dynamics of attention, apart from determining which stimuli are targets and which are not (but see Dux & Marois 2009 for a discussion of the putative role of distractor inhibition in the attentional blink).

Targets nodes which have been activated excite the transient attention node referred to here as the blaster. When this attention node crosses its threshold, the node provides a multiplicative gain across the entire set of input nodes, which boosts their ability to activate type nodes. When a type node is sufficiently activated by input, it triggers an encoding process, which takes several hundred milliseconds to complete. During encoding, recurrent feedback between binding pool nodes and the type node is established in a manner similar to the putative role of recurrence as described by Lamme & Roelfsema (2000). This feedback is important because it allows the activation of the type node to outlast the duration of the stimulus at the input layer, persisting in fact until encoding is complete. During this period of encoding, activation accrues steadily across a population of trace nodes in the binding pool and the token layer. When one of these nodes crosses threshold in the token layer, encoding is complete and the recurrent circuit between the type node, the binding pool and the token layer collapses, leaving just the stored representation of the token. This encoding process is described in greater detail below.

Critically, activation of the type node is required only during encoding into memory, and not for maintenance of that information. This detail is important because it would be difficult to encode a repetition of the same item at any point within a single trial if the type node was used to store the memory. However it is generally easy to encode and report two instances of the same target if they are sufficiently far apart in time (see Wyble et al 2009a for a discussion of repetition blindness effects and how they relate to the eSTST model).

As regards working memory capacity limits, in the following simulations, it is assumed that there are sufficient tokens to encode all of the targets presented in a trial. To simulate the inherent variance between different stimuli, targets are activated with different strength values from one trial to the next. When simulating an experiment containing multiple targets, target strength values are chosen independently and systematically for each different target. Thus, to simulate 4 targets for each of 11 different target strengths, simulations are run for each of the 14,641 (e.g. 11^4) unique combinations of these strength values.

Single target

On presentation of a single target item, a chain of events occurs over a period of approximately 400 simulated milliseconds that culminates in the encoding of the stimulus into memory as a tokenized representation. Figure 5 depicts these events as activation traces of nodes within the model starting from input at the bottom and progressing upwards to the encoding of a token. First, within an RSVP stream, depicted as a series of ovals, the target representation is activated at the input layer and excites attention. When attention crosses threshold, indicated by the symbol θ in the figure, input activations are amplified (see *a* in figure 5). Next refer to *b* where this amplified input activates that target's *type*

representation. Activation of the type node initiates encoding of a *token*. While the token is being formed in WM, attention is suppressed, producing an attentional blink as seen at *c*. (the attentional blink would affect a second target that appeared shortly after the first target but there is no second target in this example). Finally, referring to *d*, the token has reached sufficient strength and encoding is complete; the type node activation is freed to return to baseline and the suppression of attention is ended.

Lag 1 sparing: a Multi target episode

Figure 6a illustrates the encoding of two targets within a single episode. In the top panel, two targets are presented in immediate succession at an SOA of 100 ms, producing lag 1 sparing. Here, the T2 arrives soon enough to hold the attentional gate open, extending the episode to include both T1 and T2. Note that encoding of the two types overlaps. When encoding of each target is completed--first T1, then T2--activation returns to baseline. In this example, the two items were encoded in the correct order, but during sparing it can be the case that T2 completes first, producing a temporal order error, a point we return to later.

Multiple items processed within a single episode interfere weakly with one another through lateral inhibitory projections. In the case of two targets, this interference produces a slight reduction in T1 performance at lag 1 (Chun & Potter 1995; Bowman & Wyble 2007; Wyble et al 2009a), because the enhanced T2 interferes with T1 processing. This interference has a much larger influence on behavior when three or four targets are in the same episode, as will be seen below. Signs of this interference can be seen in figure 6a, as the T2 activation trace rises slightly at the end of T1 encoding.

The attentional blink: enhancing episodic distinctiveness

In the bottom panel (Figure 6b), the T2 arrives at lag 2 following an intervening distractor. Here, the T2 arrives too late to hold the attentional gate open and thus has missed the opportunity to join the episode initiated by the T1. In a more natural visual context, the T2 representation might persist in early visual areas and begin a second episode once the prior episode has been encoded. However, in RSVP, the backward masking from the following distractor wipes out the trace of the T2 so quickly that it can fail to be encoded during the delay.

What role for Capacity Limits?

Capacity limits of two types are often discussed in experimental paradigms involving rapidly presented stimuli. The first of these is a limitation on the pool of available *processing resources*, which limits the rate at which information can be encoded into memory. A classic description of this capacity limit is that of a bottleneck, and this resource limitation has been proposed as an explanation of the attentional blink (Dell Acqua, Jolicouer, Luria & Pluchino 2009, Dux, & Marois 2009).

A second putative capacity limit is on the overall amount of information that can exist within a working memory store. This limitation has been invoked to describe why performance in RSVP tasks with four or more targets is degraded for the last targets in the stream (Olivers & Meeter 2009; Nieuwenstein & Potter 2006).

Both of these limitations can reduce the overall number of items reported on a given trial. Presenting too many targets within a short time window can make it impossible to encode all of the items presented within a trial, and trying to hold too many items in a memory store at the same time can likewise limit the number of reported items (Davelaar 2005). However, in a model such as eSTST, which maintains a clear distinction between encoding of a target,

and the subsequent maintenance of that target in a memory store, it is possible to separately consider limitations on processing and storage.

With regards to processing and storage capacity limits, the eSTST model has a weak form of the former, and none of the latter. Lateral inhibition in the type layer, which can be seen in figure 4, causes each active type node to weakly suppress every other type node, thereby simulating the small but consistent cost of encoding multiple items at the same time. The interference is essential to accurately simulate the pattern of target report in the data we present below as well as competition effects described in other RSVP experiments (Potter, Staub & O'Connor 2002). However this weak interference does not cause the major portion of the attentional blink, a point we return to at length in the general discussion. As to the second type of capacity limit, the eSTST model does not simulate an overall limit on the number of targets in working memory. Close inspection of the data from the present experiments does not suggest the involvement of this type of capacity limit in these experiments, a point we discuss in the final prediction of the paper.

Experiments and Predictions

Here, the eSTST model is used to generate seven predictions that explore the boundary conditions of what constitutes the break between attentional episodes and also the costs and benefits of allowing multiple items access to working memory during a single attentional episode. These predictions are provided by simulating new experimental conditions with the *same set of parameters* used in the original publication of the model (Wyble et al 2009), with only the addition of new type and token nodes to allow up to 6 targets to be presented to the model. Thus, these simulations do not represent "fits" of the model in the usual sense, because no parameter adjustments were permitted to fine tune the model to the data at hand, apart from a change to reflect the stronger physical masking used in Experiment 3. Three main experiments are used to evaluate these predictions.

Experiment 1

What are the benefits and costs of encoding multiple items that are selected for working memory encoding during a single attentional episode? In this experiment, subjects viewed four targets in different configurations.

Predictions for Experiment 1

Prediction 1. At high rates of presentation, more targets can be remembered when they are clustered than when they are spread out—The model predicts that when targets are presented at a rate of 100 ms/item in immediate succession (400ms total time), they keep attention engaged, and thus, clustered targets are more frequently encoded than when those same targets are alternated with distractors (700ms total time).

Prediction 2. Temporal spacing determines episodic continuity—The model predicts that it is the temporal spacing of new target items that is most critical in sustaining an attentional episode. If targets are presented within about 100 ms of each other, attention can be sustained even during the presence of intervening distractors.

Prediction 3. Episodic segmentation enhances temporal order information—In the model, interleaving a distractor between two targets delays the encoding of the latter target, and thereby increases the accuracy of reporting the sequential order of the two targets.

Methods

Participants—The ten participants were volunteers from the MIT community between the ages of 18–35 who were paid to participate in the experiment, which took approximately 30 minutes. All reported corrected or normal vision.

Apparatus and stimuli—The experiment was programmed using Matlab 5.2.1 and the Psychological Toolbox extension (Brainard, 1997), and was run on a PowerMac G3. The Apple 17" monitor was set to a 1024 × 768 resolution with a 75 Hz refresh rate. An RSVP stream was presented centrally at the location of a fixation cross.

Black digits (2,3,4,5,6,7,8,9) in 70 point Arial were used as distractors. Capital letters (B, C, D, F, G, H, J, K, L, N, P, Q, R, T, V, X Y, Z) were used as targets. Stimuli were approximately 1.5 by 1 degrees in angle at a viewing distance of 50 cm. These stimuli comprised RSVP streams presented at either 53 or 107 ms per item with no interstimulus interval.

Design and procedure—Each trial consisted of an RSVP stream, which included four single-letter targets (none repeated) among digit distractors. In the 2 × 2 design, one factor was successive (TTTT) or separated targets (TDTDTDT); the second factor was the presentation rate, 53ms or 107 ms per item. There were two blocks of 120 randomly mixed trial types.

Each trial began with a fixation cross for 1 second and a sequence of 7 to 12 distractors on the slow (107 ms) trials and double that number on the fast (53 ms) trials, so that the average time before the first target was equated for the slow and fast RSVP rates. At least 5 distractors followed the last target. Subjects were instructed that they would see four letters and should remember them for entry at the end of the trial. They were told that they were free to report the targets if they were not sure, but not to guess randomly. Report of temporal order of the targets was implicit in the response prompt provided to subjects, which appeared in a sequence of characters from left to right as subjects type in their response. We intentionally avoided giving explicit emphasis to order information in the instructions out of concern that such emphasis would come at the expense of item information. Subjects were allowed to correct their input string with backspace while entering it, and were given feedback as to the letters they saw and their correct order. Responses were considered correct if subjects reported the correct identity, without regard to correct order although we did analyze the pattern of order errors for reported targets.

Model Simulation

The predictions of the eSTST model for each of the four conditions were calculated. As the model uses time steps of 10 ms, presentation was simulated at 50 ms and 100 ms per item rates. In the simulation, the input strengths of the targets are varied over trials and performance is averaged over these trials to derive behavioral accuracy curves. This strength value represents the processing difficulty of each target, which is assumed to vary due to differences in backward masking strength produced by the interaction of each target and the immediately following item, as well as intrinsic differences in the processing of each target, due to such factors as familiarity and orthographical or phonological similarity with other members of the target set, etc.

Figure 7 shows traces of the model's performance in the four conditions for just one particular item strength (i.e., one trial). The model does not behave in the same way on every trial due to differences in item strength, but in the figure we have used a strength value

in the middle of the input range (.9) for all four targets to elicit broadly representative activation traces.

Simulated trial: 100 ms per item—As previously described in Wyble et al. (2009a), presentation of four targets in direct succession (TTTT) at 100 ms per item produces a protracted sparing effect, simulating the pattern observed in Olivers et al. (2007) and Kawahara et al. (2006). Attention triggered by the T1 spills over to the following item and amplifies the representation of T2, which in turn sustains the level of attention. This dynamic continues for T3 and T4, although the accumulating interference between simultaneously active type nodes produces progressively weaker encoding. In Figure 7, for the trial in the upper left hand corner, all four targets are correctly encoded, and in the correct order.

When targets of the same input strength are presented for the same 100 ms duration, but are now separated by distractors (TDTDTDTD), the dynamics of attention are remarkably different. Here, under the suppression from T1 encoding, the attentional gate is closed during the interval between the first two targets. Attention is thus suppressed when T2 arrives and remains so until T1 encoding is completed. In the illustrated trial, T1 encoding is complete after approximately 300 ms, freeing up the attentional gate in time for T3 to be encoded, which produces another period of attentional suppression that keeps the T4 from being encoded. In this example trial (bottom left panel of Figure 7), attention has segmented the input into two episodes; successfully encoding T1 and T3 without overlap, but at the cost of missing T2 and T4 entirely. This episodic division is evidenced by the two peaks of attention in the simulated TDTDTDT @ 100ms SOA trial in the lower left corner of Figure 7. Note that on other trials, using a different set of strength values for the targets, the predicted pattern of which targets are encoded would be different than this simulation.

Simulated trial: 50 ms per item—Simulating the results at a faster rate of 50 ms per item, TTTT and TDTDTDT both produce a sustained attentional episode in the right panel of Figure 7. In the simulated TTTT trial, all four targets are encoded, although in this particular example, the order is incorrect: T2,T3,T1,T4 as can be seen in the relative times at which the type node activations return to baseline. For TDTDTDT trials, even though there are intervening distractors, the T2 arrives rapidly enough to benefit from the attention elicited by T1, and bolsters the deployment of attention against the suppression produced by encoding. In this example trial (but not in all cases), the building interference from the ongoing encoding of T1,T2 and T3 results in the episode being concluded prematurely and T4 is lost, producing the encoded sequence: T2, T1, T3. In other trials, a stronger input strength for T4 would allow it to be encoded along with the preceding targets.

Simulated blocks of trials—An experimental block that uses randomly selected targets and distractors can be simulated by averaging over single trials that vary in their strength values. For the following simulations, the strength of input values covered the same range as in Wyble et al (2009a). Strength of an individual target varied in steps of .1080 over the range .31 to 1.39 for a total of 14,641 simulated trials (i.e. all 11^4 combinations of target strength for the four targets).

The overall pattern of simulated accuracy in the four experimental conditions is depicted in Figure 8a–b, alongside the results of human subjects performing the same conditions as in Figure 8c–d. In the human data, an ANOVA using position (T1-T4), rate of presentation (53 ms or 107 ms) and trial structure (TTTT vs TDTDTDT) as factors found all three main effects and four interactions to be significant at least to the .01 level (all F's > 10 and $\eta_p^2 > .54$). Focussed analyses address each of the predictions.

Prediction 1: At high rates of presentation, more targets can be remembered when they are clustered than when they are spread out—This ability to process multiple targets simultaneously with only a modest amount of interference is central to eSTST. As can be seen in Figures 3, 6, and 7, in the model, multiple types are concurrently encoded over several hundred milliseconds. Because attention is strongly engaged by a string of targets, performance is actually better when targets are presented closely together in time. This is visible in the simulations shown in Figure 7 by comparing the top left panel to the bottom left panel. The traces indicate that all four targets are encoded in the TTTT condition, but not in the TDTDTDT condition at 100ms SOA.

The human data are entirely consistent with an enhanced ability to process multiple targets arriving in immediate succession. A highly reliable finding is that average accuracy was better for targets presented in a temporal cluster as compared to targets distributed over a longer time span. For the 50 ms SOA condition, TTTT were presented over a 200 ms window and TDTDTDT were presented over 350 ms. Overall performance was improved in the TTTT compared to the TDTDTDT condition (52% vs 48%, $F(1,9) = 8.3$, $\eta_p^2 = .48$, $p < .02$). In the 100 ms SOA condition, the condensed target presentation (TTTT) produced a more marked improvement in accuracy compared to the TDTDTDT condition (62% vs 52%, $F(1,9) = 32.2$, $\eta_p^2 = .78$, $p < .001$). As subjects are doing better on trials in which they are given less total time to process targets, it seems clear that resource limitations cannot be the primary cause of encoding failures in these trials. There is, however, a potential confound of subject expectancy effects produced by mixing the slow and fast trials together and we address this problem in Experiment 1a below. It is also notable that despite overall enhanced performance for the four targets, accuracy of the first target is slightly lower in the TTTT condition than in the TDTDTDT condition. This is expected based on the usual finding that T1 accuracy is typically reduced when a second target follows it by 100ms or less (Potter Staub & O'Connor 2002). The eSTST model also simulates this small difference in T1 accuracy between TTTT and TDTDTDT as a consequence of the weak competition between coactive types.

Prediction 2: Temporal spacing between targets determines episodic continuity—In the simulation, the presence of interleaved distractors at the 50 ms presentation rate does not produce an episodic division. This comparison can be seen in the simulations shown in Figure 7 by comparing the top left panel to the bottom right panel. In both cases there is a single attentional window, as can be seen by the trace of attentional activation. To test this prediction, we compare performance between TTTT at the slow rate and TDTDTDT at the fast rate. In both of these conditions, targets appear at 107 ms intervals, so the temporal arrangement of targets is preserved but the presence of intervening distractors is varied. Figure 9 depicts the comparison between these conditions in the human data. A planned comparison in the experimental data confirms this observation, showing main effects of target position ($F(3,27)= 21.5$, $\eta_p^2 =.71$, $p < .001$), and condition ($F(1,9) = 78$, $\eta_p^2 =.90$ $p < .001$), but no interaction ($F(3,27) = 1.7$, $\eta_p^2 =.15$, $p > .19$).

Prediction 3: Episodic Segmentation enhances temporal order information—
The model predicts that temporal information will be enhanced for two targets if attention segments them into separate episodes. Specifically, on trials in which two given targets were successfully encoded, their reported order will be more accurate if those targets were separated by a distractor, while holding the temporal interval between them constant. The comparison best suited to address this question contrasts T1 and T3 in the slow TTTT condition, with T1 and T2 in the slow TDTDTDT condition. In both cases, the TOA between the targets of interest is ~200 ms and the targets have been presented for the same total duration. Figure 10 illustrates why the model has greater difficulty encoding order correctly when presented with TTT than TDT; in the former case, all three targets are

encoded concurrently while in the latter, encoding of the second target begins when encoding of the first target is complete. The strength values of the targets are the same in the two simulations, but the middle target of TTT allows attention to be sustained, creating an episode that includes all three targets.

Order information can be measured in terms of "correctly ordered response pairs" (Reeves & Sperling, 1986), which computes, for all trials in which items *i* and *j* are reported, the probability that they are reported in the correct order. Using this metric, we computed the model's p(correct order|T1^T3) in the TTTT condition and p(correct order|T1^T2) in the TDTDTDT condition (i.e. considering trials in which the relevant targets were correctly reported and ignoring performance on the other targets). In the simulation, when the two relevant targets were encoded during the same episode (TTTT) they were correctly ordered only 85% of the time, in contrast to perfect ordering (100% order accuracy) in the TDTDTDT condition. In the human data, p(correct order|T1^T3) in TTTT was .79 and p(correct order|T1^T2) in TDTDTDT was .91 (paired t(9) = 3.07, one tailed *p* < .01, Cohen's *d* = .35). Thus, this prediction is confirmed, with order information for temporally equivalent targets in TTTT being significantly worse than TDTDTDT.

Experiment 1a

In Experiment 1, in 75% of the conditions all four targets arrived within a window of 428 ms or less. To ensure that poor performance on slow TDTDTDT trials was not the result of an expectation on the part of subjects that the targets will arrive in a short temporal window another experiment included only slow (107 ms) trials, presented with either a TTTT or TDTDTDT structure. In this replication, ten participants saw an equal number of trials in condensed and distributed presentations in two mixed blocks of 120 trials. The results replicated the data of Experiment 1 for the slow trials in every respect as shown in Figure 11. As before, condensed target presentation produced superior report accuracy relative to distributed targets (65% vs 54%, F(1,9) = 30.6, η_p^2 = .78, p < .001). This experiment also replicated the ordering accuracy difference between p(correct order|T1^T3) in TTTT, which was .79 and p(correct order|T1^T2) in TDTDTDT, which was .89 (paired t(9) =3.76, one tailed *p* < .005, Cohen's *d* = .31).

Experiment 2, Two Episodes

How readily can subjects encode two episodes and how does encoding of a prior target affect encoding of the current one? In this experiment, we asked subjects to report 6 targets. As before, in one condition, the targets were interleaved with single distractors (TDTDTDTDTDT). In the other condition, targets were presented in two clusters of three, *but over the same total time interval* (TTTDDDDDTTT). All items were presented for 107ms. If the model's implementation of attention is accurate, subjects should perform better in the latter condition than the former. Again, the simulations were run with the original parameters of the model.

Predictions for Experiment 2

Prediction *4. Subjects can efficiently encode two multi-target episodes*—In clustered presentation, performance averaged over both clusters of targets should be superior to the interleaved condition; subjects should be capable of encoding two episodes provided they are separated by an interval that is sufficiently long to allow the encoding of items acquired during the first episode to be completed.

Prediction *5. Inter-target contingencies are altered by clustered presentation* **—**The model predicts that attention functions differently during clustered vs interleaved

target presentation; with clusters, reporting target T_{n-1} will only weakly affect report of target T_n, because the close temporal proximity allows attention to be sustained across an entire episode. With interleaved distractors, successful report of each target T_{n-1} will have a potent detrimental effect on the report of the following item T_n, even though the two targets are now further apart in time, because successful encoding of T_{n-1} will have suppressed attention to T_n.

Method

The method was similar to that of Experiment 1 with the exceptions described below.

Participants

Fourteen participants were drawn from the same subject pool as that of Experiment 1.

Design and Procedure

All stimuli were presented for 107 ms. There were two trial types, intermixed randomly, in two identical blocks of 120 trials. Either the six targets were separated by distractors (TDTDTDTDT) or appeared in two clusters (TTTDDDDDTTT). In each case, the temporal interval between the onset of the first and last targets was 1067 ms (in the simulation, 1000 ms). As in Experiment 1, the target sequences were preceded and followed by additional distractors.

Model Simulation: TTTDDDDDTTT vs. TDTDTDTDT

In the 6 target simulations, strength of an individual target varied in steps of .27 over the range .31 to 1.39 for a total of 15,625 simulated trials (i.e. all 5^6 combinations of target strength for the six targets). To compensate for the large step size, random variance was added to each target strength value and was chosen from a uniform distribution of the range $[-.09\ .09]$.

In the results of the model simulation (Figure 12), report accuracy is superior for the clustered target presentation. Within each cluster, the close spacing of targets sustains the deployment of attention and targets are well encoded. The separation between the clusters permits an episodic break, which allows processing of the first episode to be completed before the second begins, producing excellent performance for all six items. When six targets are evenly distributed over the same temporal interval (1,100 ms), the simulated pattern of accuracy is distinctly different; performance decreases sharply for the second target and remains well below T1 levels until the end of the target sequence. As with Experiment 1, the 200 ms TOA between targets does not permit an attentional episode to be sustained and the attentional gate is intermittently opened and closed producing an overall reduction in average performance across trials.

Results

As shown in Figure 12, participants in the same two conditions gave results similar to the model's with the exception that T4 accuracy (the first target of the second cluster) is quite low in the two cluster condition. This pattern, consisting of poor accuracy for one target that gives rise to much better performance on a subsequent target is reminiscent of cueing effects within the AB (Nieuwenstein, Chun, van der Lubbe, & Hooge, 2005) and suggests that the T4 is acting as a cue for the following target in the second episode.

This result suggests that processing of the first cluster of targets is protracted and the T4, arriving 533 ms after the T3, is still within an attentional blink induced by the first episode. In the model, a parameter corresponding to the rate of WM encoding determines the

duration of the blink and this parameter value does not capture the full extent of the blink duration in this case.

Prediction 4: Subjects can encode two multi-target episodes

As in Experiments 1 and 1a, overall accuracy in the clustered presentation is substantially superior to that of the interleaved target presentation. An ANOVA using position (T1-T6) and trial structure (TTTDDDDDTTT vs TDTDTDTDTDT) as factors, found target position to be significant ($F(5,65) = 51.7$, $\eta_p^2 = .79$, $p < .001$). Furthermore, the targets presented closely in time were more accurately reported than distributed targets (66% vs 54%, $F(1,13) = 95.2$, $\eta_p^2 = .88$, $p < .001$). The interaction between these variables was also significant ($F(5,65) = 35.4$, $\eta_p^2 = .80$, $p < .001$).

Prediction 5: Inter-target contingencies are altered by clustered presentation

In the eSTST model, the sustained reduction in performance for T2-T6 in the interleaved condition is essentially the superposition of multiple attentional blinks at different lags in different trials. For example, on one trial, T1, T4 and T6 may be encoded, while on another trial, T1, T3 and T5 may be encoded. In contrast, when targets are presented more closely in time (i.e., with an TOA of 100 ms), each encoded target weakly competes with the following target, but also helps to sustain attention, and the sum of these two effects washes out so that there is a minimal impact of encoding one target on encoding the next.

This difference can be quantified by comparing performance on targets T2 through T6 as a function of whether the preceding target was seen or missed: $p(T_n|T_{n-1})$ and $p(T_n|!T_{n-1})$. These paired measurements are shown for simulations and human data in Figure 13. An ANOVA on the human results was performed with the factors of target position (T2-T6; $F(4,52) = 28$, $p < .001$, $\eta_p^2 > .68$), trial structure (TTTDDDDDTTT vs TDTDTDTDTDT; $F(1,13) = 197$, $p < .001$, $\eta_p^2 > .94$) and prior target report (T_{n-1} seen vs missed; $F(1,13)$, $p < .001$, $\eta_p^2 > .54$). The two interactions: target position X trial structure ($F(4,52) = 16.3$, $p < .001$, $\eta_p^2 > .55$) and trial structure X prior target report ($F(1,13) = 20.5$, $p < .001$, $\eta_p^2 > .61$), were significant. The three way interaction was not significant. This pattern of effects suggests that prior target encoding affected accuracy differently in the two trial structure conditions, but this effect was not particular to specific target positions (T2-T6) as the target position X prior target report interaction was not significant ($p > .25$). Therefore the impairment due to seeing the T_{n-1} target was greater in the interleaved condition and this impairment was not dependent on target position.

This analysis shows that the model correctly predicts the dynamics of encoding targets in the two presentations. In the interleaved condition, perception of each target impairs report of the following item, and this does not occur as strongly when the targets are clustered together. Note that the relatively small T_{n-1} contingency effect reported here for clustered targets (TTT) is similar to the Within Trial Contingency effect (Dell Acqua, Jolicoeur, Luria & Pluchino 2009; but see Olivers Hulleman, Spalek Kawahara & Di Lollo in press). This reflects weak interference between multiple items within an episode, an effect that we see as distinct from the attentional blink. A supplemental section, available online, illustrates the results of conditional analyses of the data from experiments 1 2 and 3 for trials in which T1 was reported correctly, alongside simulations of those conditional analyses.

Experiment 2a: Increasing the separation between episodes

In the data of Experiment 2, subjects had difficulty reporting the first target of the second episode in the TTTDDDDDTTT condition. This suggests that the gap between the two episodes was not long enough, and that subjects were still encoding the first episode when the second episode began. As the two episodes were 525 ms apart, this explanation would

imply that attentional blinks produced by episodes containing several targets last considerably longer than blinks produced by single targets. There is support for this idea; an experiment by Ouimet & Jolicoeur (2006) demonstrated that encoding a T1 composed of multiple simultaneously presented digits produces an AB that requires more than 1200 ms to fully recover. To test whether prolonged encoding time for the first episode was responsible for the drop in performance on the first target in the second episode, in Experiment 2a, we extended the interval between the episodes by adding three more distractors, producing the stream TTTDDDDDDDDDTTT. In the interleaved case, the new distractors were interspersed between the targets as follows: TDDTDTDDTDTDDT, producing the same total time from first target to last target.

The results of this experiment, shown in Figure 14 alongside the simulated equivalent, agree with the prediction that some of the impaired accuracy of the first target in the second episode in Experiment 2 was a result of an interval between the episodes that was too short. In this experiment, encoding of the two target clusters is more similar, and overall performance remains markedly superior for the clustered targets than the distributed targets (63% vs 49%, $F(1,9) = 36.8$, $\eta_p^2 = 81$, $p < .01$). However, the second cluster had slightly lower overall accuracy than the first, as shown in a focused analysis that considered only TTTDDDDDDDDDTTT trials, with a single factor that compared overall accuracy in the first and second cluster (67% vs 59%, $F(1,9) = 10.9$, $\eta_p^2 = .55$, $p < .01$). Finding this difference in accuracy despite the fact that the blink induced by the first episode had recovered suggests the influence of working memory capacity, an issue we focus on in the following section.

Another obvious effect in this new experiment is the sawtooth pattern of accuracy in the TDDTDTDDTDTDDT trials, both in the simulation and the empirical results. The targets which are the least often reported, in both data and simulation, arrived just after the double distractors (positions 2, 4 and 6). This pattern suggests that the longer gap between targets allows greater suppression of attention, thus reducing performance on the following target.

Experiment 3

This final experiment explores the boundary condition of the termination of an attentional episode. It has been suggested that the dynamics of attention are primarily temporal in nature, rather than being driven by the sequential presentation of targets and distractors (Nieuwenhuis et al 2005; Bowman & Wyble 2007; Wyble et al 2009) This idea has been buttressed by findings which show that the course of the attentional blink is anchored to the time of the T1 presentation, rather than the number of distractor items following it (Bowman & Wyble 2007; Martens, Munneke, Smid & Johnson 2006). The eSTST model is likewise temporal, and it predicts that the continuation of an attentional episode is determined by the temporal continuity of target spacing. Distractors can play a helpful role in delineating the end of an episode, but they should not be necessary. In this experiment a T4 is presented following a cluster of three targets, and they are separated by a blank gap rather than by distractors. These results demonstrate that episodic changes in attention occur in the absence of distractors before and after the cluster of targets. This experiment extends the findings of earlier research (Nieuwenstein, Potter & Theeuwes 2009) illustrating an attentional blink following a gap, but in this case the stimuli that elicit the blink are a cluster of three targets and the T4 is presented for 107ms, a more conventional SOA in RSVP paradigms, rather than the 58ms duration used in the prior research.

Prediction 6. An attentional blink can be evoked by a temporal gap at the end of a sequence of successive targets

According to the eSTST model, the necessary condition for producing an attentional blink is a sufficiently long temporal gap between the final target in a sequence and the next target. However, this blink will be substantially weaker in magnitude if no intervening distractors are present (see experiment 3 of Olivers et al (2007)). The eSTST model (see also Bowman & Wyble 2007; Bowman et al 2008) explains the role of the post target mask as extending the duration of target processing for stimuli that are presented briefly (i.e. about 100ms), and thereby increasing the depth and duration of the attentional blink.

Method

The method was similar to that of Experiment 1 with the exceptions described below. Four letters were presented one after the other at the center of the screen for 107ms each, without distractor items. A backward mask was shown after the fourth letter. Participants saw a fixation cross for 1000 milliseconds, followed at a randomly chosen interval from 856ms to 1284 ms by three sequential letters for 107 ms each. The fourth letter was presented from 1 to 7 positions after the first three targets as illustrated in Figure 15 and was followed by a mask composed of an @ symbol superimposed on top of a # symbol for 107ms. This mask was used because it is effective as a trailing mask, and also because it is not an easily reportable character that subjects might inadvertently encode as a potential target. No other stimuli were presented, until the response screen appeared 535 ms after the fourth letter.

Model Simulation: TTT_____TM at different lags

Figure 16 illustrates the predictions of the model when T4 is presented at various lags from T3 without any intervening distractors. Note that in the simulation, T1 performance is better than T2 performance, unlike previous simulations in this paper. The reason for this difference, as explained by the eSTST model (Wyble et al 2009), is that Experiment 3 has no distractors prior to T1. Therefore the participant can process targets unselectively, producing the same declining pattern of accuracy as found in whole report (Nieuwenstein & Potter 2006). To simulate unselective processing in the model, the delay of attentional deployment is reduced from 40ms to 10ms (see Wyble et al 2009), which gives the T1 a competitive advantage over T2 rather than vice versa.

In the simulation, T4 performance remains relatively lower than T1, even at lag 7 after recovery from the blink. This is due to simulation of the more potent perceptual mask presented after T4 compared with the T1 which was masked by another letter (i.e. the T2). We simulate this enhanced masking by reducing the overall strength of the T4 representation.

Results

The results of the experiment are shown in Figure 16, showing T1-T4 performance for the 7 lag conditions of T4. T1 and T2 did not differ in any systematic or predicted way across the 7 lag conditions. T1 performance ranged from 83% to 89%. T2 performance ranged from 65% to 75%. Thus the illustrated T1 and T2 scores are averaged across lags 1–7 and these accuracies are 85% and 70%. T3 performance differed markedly between lag 1 (66%) and lags 2–7 (82% to 92%) as predicted by the model. This is expected given that at lag1, T3 was masked by T4, but at longer lags it was unmasked. T3 performance was significantly worse when T4 was at lag1 (66%) than lag 2 (82%) (paired t (11)= 3.5, p < .003 one tailed, Cohen's d = .44).

For T4 accuracy, there was a significant effect of lag (F(6,66) = 5.1, η_p^2 = .32, p < .001). A focused test found a predicted decrement in accuracy between lag 1 (51%) and lag 2 (42%) (paired t(11) = 2.5, p < .05 one tailed, Cohen's d = .21). There was also a predicted improvement in performance between lag 2 and lag 7 (paired t(11) = 3.9, p < .0015, one tailed, Cohen's d = .32). Thus, we obtained an attentional blink at lag 2 that recovered gradually over the course of hundreds of milliseconds despite the lack of distractors between the preceding targets and the blinked target.

Discussion of Prediction 6

These results chart the onset and recovery of a blink following a three target episode in the absence of any distractors, apart from the trailing mask of the T4 (which is necessary to observe the blink). This result supports the findings of Nieuwenstein, et al. (2009; see also Nieuwenstein, Van der Burg, Theeuwes, Wyble, & Potter, 2009) & Visser (2007) in which an attentional blink was found in experiments involving just two targets separated by a blank temporal gap.

The eSTST model predicts that a blank temporal gap initiates an attentional blink since it provides a period of time during which inhibitory control (c.f. Figure 2) has an opportunity to win the competition for control of attention. As a result, T1, T2 and T3 enter the encoding process as a single episode, and T4 encoding is delayed while the first three targets are encoded. Because T4 is strongly masked, the delay of encoding at short lags results in lower accuracy. Another important facet of this result is that we observed an attentional blink following a blank gap for a T4 that was presented for 107ms. In Nieuwenstein et al (2009), an attentional blink was not observed for a 100ms T2 following a single target. The model suggests that this effect is obtained because a cluster of three targets are encoded simultaneously, producing a lengthier suppression of attention than does a single target. The fact that three targets produce a measureable blink for a 107ms target lends further support to the theory that multiple targets presented in a single cluster are encoded simultaneously.

Interference within an episode vs overall capacity limitations

In RSVP target sequences containing three or more sequential targets followed by distractors, performance begins to degrade as the sequence progresses. This pattern is illustrated in figures 8, 9, 11, 12, and 14 of this paper, as well as the data of Olivers et al (2007), Kawahara et al (2007), Nieuwenstein & Potter (2006) and Wyble et al (2009a). Other contemporary accounts of RSVP performance, such as Nieuwenstein & Potter (2006) and Olivers & Meeter (2009) describe the cause of this drop in performance over a string of to-be-remembered items as a memory limitation. However, the eSTST model offers a different explanation: as additional targets are added to an episode, earlier targets are still being encoded, which slightly reduces the probability of encoding the new targets. This encoding difficulty stems from two sources.

1. *Weak interference*

During encoding, simultaneously active items weakly interfere with one another directly at the type layer (the lateral inhibitory connections in the *Types* layer in Figure 4). This simulated interference is needed to explain why T1 is reduced in accuracy when T2 directly follows it at lags of 100ms to 150ms during selective report from an RSVP stream (Chun & Potter 1995; Potter, Staub & O'Connor 2002) as well as the weak T_n inter target contingency effect found in prediction 5. This interference effect is weak however, and is not capable of causing the attentional blink.

2. *Inhibitory control*

Inhibition of attention due to encoding (the inhibition of the blaster in Figure 4) grows stronger as more items are being encoded. This inhibition can prevent a particularly weak target from keeping the attentional gate open, thereby reducing the proportion of targets which are encoded in positions 3 and 4 of an episode.

Both of these effects reduce performance on later targets within an episode and these effects are relieved when there is a gap between targets that allows the encoding process to run to completion. This facet of the model leads to a specific prediction that we evaluate by revisiting data from experiments 1, 2, and 3.

Prediction 7. Recovery of encoding capacity with time

In eSTST, the drop in performance for T4 in the sequence TTTT is caused by short-term interference within an episode that recovers over time.

To evaluate this prediction, we consider the comparison between performance in the three conditions illustrated in Table 1, all of which are different arrangements of targets presented at 107ms SOA. Consider the relatively low performance (43%) of T4 in TTTT of experiment 1. If this reduced performance is indeed due to an overall working memory capacity limit as suggested by Olivers & Meeter (2008), performance of all targets arriving later should be at this level or worse. Contrary to this prediction, performance for T5 is 67% in both of the six target conditions and T6 is reported at 65% and 58%. In the eSTST model, this performance recovery is due to the temporal gap between T3 and T5, which allows encoding of T1,T2 and T3 to complete. Thus, it seems that for a string of targets presented at RSVP speed, there is an accruing interference effect that is better captured by the eSTST model than by a working memory storage capacity limit.

For another example of this recovery, consider the results of Experiment 3. In accord with the prediction of the eSTST model, performance on the T4 at lag 1 is worse than T4 performance at lag 7. Thus, the impairment of T4 at lag 1 could not have been due solely to a hard limit on working memory. In fact, at lag 7, not only is T4 performance improved relative to lag1, but T3 performance is markedly improved as well, which contradicts the explanation offered by overall working memory capacity.

It should be emphasized that these experiments do demonstrate some form of capacity limitation. This is most clearly evident in experiment 2a in which the second episode has apparently escaped the blink, yet its overall accuracy is significantly worse than report of the first episode. The eSTST model, which has no capacity limit, fails to simulate this difference. Adding a capacity limit to the model is a potential avenue for exploration, but it is not yet clear how to represent such capacity. Previous work with a binding pool mechanism in a type/token framework is suggestive of a distributed form of capacity that degrades gracefully under increasing load (Wyble & Bowman 2006) but additional data is necessary to constrain such a model.

General Discussion

Is visual attention episodic in the way described by the eSTST model? The present study confirms key predictions of the model, which suggests that the competitive interplay between working memory encoding and attentional selection results in a visual mechanism that is responsive to the temporal structure of its input. In particular, the data demonstrate that participants are able to report more RSVP targets when presented in clusters, and the model suggests that this mechanism serves to encode temporally proximal information within a single episode. In the model, a temporal gap between two targets of 200 ms or

longer produces an effect whereby the later target is encoded in a subsequent episode, and the consequent suppression of attention produces an attentional blink. The ability of a stimulus to enter working memory is significantly compromised during this period, especially if the stimulus is briefly presented. The duration of this window of suppression can be sufficiently long (e.g. over 500 ms in Experiment 2; and see also Ouimet & Jolicoeuer, 2007) to explain well known failures of perception in situations such as stage magic, pick-pocketing, or information- rich environments such as cockpits (Su, Bowman, Barnard, & Wyble 2009). In all of these cases, apparent lapses in attention may occur not just from spatial capture to an inappropriate location, but from the temporal structure of events creating periods of inattention.

These predictions illustrate several properties of attentional episodes. First, for all five of the experiments overall performance is superior for targets presented in clusters in comparison to conditions of interleaved distractors. Next, prediction three illustrates that while targets presented in clusters are reported more often, this enhanced report comes at a cost of temporal order information, even when TOA is held constant.

Experiments 2 and 2a demonstrated that two clusters can be encoded within a single trial, each of which has a similar pattern of performance that peaks at its second target. Furthermore, performance for a given target within a cluster is not strongly affected by successful report of a previous target, unlike the case when targets are separated by distractors.

The boundary condition defining the end of an episode seems to hinge on a 200ms temporal gap between target onsets rather than the presence of post-target distractors. This is suggested by two findings. First, in prediction two, it was found that distractors were insufficient to produce an AB if they were too brief. Second, in prediction six, an attentional blink is observed without the presence of intervening distractors. Therefore the data exhibit an attentional blink without the presence of post target distractors (cf. Nieuwenstein et al., 2009), and post target distractors do not necessarily cause an attentional blink.

Attentional selection and limited resources both play a role in determing performance

A prominent debate in the attention literature concerns whether limited cognitive resources (Arnell & Jolicoeur 1999, Dehaene, Sergent & Changeux 2003, Dell Acqua et al. 2009, Dux & Marois 2009) or processes of attentional selection (Olivers & Meeter 2008, Wyble et al 2009, Taatgen, Juvina, Schipper, Borst, & Martens 2009, Nieuwenhuis et al. 2005) are primarily responsible for producing the attentional blink. This question helps to define the limitations and capabilities of our ability to perceive multiple stimuli in rapid succession. The resource theory proposes that the attentional blink is the result of a depletion of central processing resources by the T1, which requires some amount of time to recover (Dux & Marois 2009). On the other hand, selection based accounts, such as the temporary loss of control (Di Lollo et al. 2005), Boost/Bounce model (Olivers & Meeter 2009), eSTST (Wyble et al 2009) and Threaded Cognition (Taatgen et al. 2009), propose that while there are sufficient resources to process multiple targets, attentional mechanisms regulate the flow of information from early visual representations into working memory, producing the attentional blink. These theories share a reliance on attentional control circuitry, but differ markedly in regards to the functional specifications of temporal attention (for review see: Martens & Wyble 2010).

The eSTST model addresses this debate by illustrating how both attentional selection and limited resources interact within the same framework. A weak form of interference (i.e. the hallmark of limited resources) occurs between multiple targets within a single episode due to their close spacing while attentional selection provides the separation between episodes.

These effects are produced by distinct mechanisms within the model. The necessity of including both selection and resource limitations in a model of performance in visual perception tasks that involve the attentional blink is in line with a broad perspective of the attentional blink literature (Kawahara, et al 2006; Dux & Marois 2009). Furthermore, it is clearly the case that there is a limit on the rate at which tokenized representations of stimuli can be perceived, even when identification is not required (Garner 1951).

The evidence for interference between closely spaced targets can be found in the diminished report of T1 during lag-1 sparing (Chun & Potter 1995), the within-trial-contingency effects reported by Dell Acqua, et al. (2009), and the effects of modulating encoding emphasis (Dux, Asplund & Marois, 2009). The model captures this interference between neighboring items with the inhibitory connections between Type nodes (see figure 4). This weak interference is critical for reproducing the pattern of declining report accuracy for targets 3 and 4 (Figures 8, 9, 11, 12, and 14). As a further demonstration of this weak interference, a supplemental section is available online, which illustrates anayses of experiments 1, 2 and 3 conditional on report of T1 alongside simulations generated from the model.

In the eSTST simulations, this weak interference is not capable of reproducing the dramatic attentional blink effect, nor can it explain why overall report of targets is higher in the TTTT condition than the TDTDTDTD condition (Experiments 1 and 1a). To explain these effects, the model requires an attentional selection component. So while the simulation of limited resources do play an important role in allowing the model to replicate the complete pattern of data as described in the preceding paragraph, such limitations are not the cause of the attentional blink. Prediction 5 (see Figure 13) makes this point explicit by showing that the weak pattern of inter-target interference present in the TTTDDDDDTTT condition is replaced by a much larger inter-target deficit when those targets are interleaved with distractors in the TDTDTDTDTDT condition. In the latter condition, report of each item is strongly diminished by the suppression of attention whenever the prior target is successfully encoded. In agreement with these results, another experiment has demonstrated that inter-target interference can be dissociated from the effects of attentional selection by varying the SOA between two targets (Olivers, et al. In Press).

Multiple targets within a single epsiode

Episodes, as simulated by the eSTST model, are not a memory structure. Rather, episodes refer to the temporal windows during which information is admitted for further processing and storage into memory. Thus, when multiple items are successfully admitted during a single episode, they are not combined into a single representation, but instead form a series of sequentially organized representations of the individual target items. The stored sequence, as noted, is sometimes in a different order than the input order. This is an important modification of the original STST model (Bowman & Wyble 2007), for which two items stored at lag-1 sparing were usually combined into a single representation that contained no temporal information.

This modification of the original STST theory is suggested by two lines of evidence First of all, multiple targets in RSVP sequences presented for 50–100 ms per item, such as those reported here, are visibly distinct as individual, sequential items according to subjects' reports. In fact, subjects frequently report a vivid awareness of the sequence order of two items presented within even a 50ms SOA, and this is true even on trials in which their report of the order is inaccurate (Caldwell-Harris & Morris 2008). A second source of evidence is that subjects do recover a significant amount of order information from an uninterrupted four target sequence. Figure 17 is reprinted from Wyble, et al (2009a), and illustrates both simulated results and empirical data from the temporal order of 4 letter targets in an RSVP stream presented as TTTT. These data represent the set of trials in which both the model and

subjects reported all four items correctly. Clearly, some, but not all, of the order information is preserved for uninterrupted target sequences. The overall pattern resembles perturbation (Estes 1997) in which order report exhibits a tendency for individual targets to switch positions with their immediate neighbors. Targets on either end of the episode are correctly positioned more often than targets in between the endpoints.

Attentional episodes, transient attention and attentional capture

Research exploring the spatial and temporal aspects of attention typically find that presenting a salient cue produces a brief enhancement of processing at that particular location that is time locked to the onset of the cue. It is possible that this spatiotemporal form of attentional deployment is mediated by the same episodic attentional control as we describe here for RSVP experiments with no spatial component (cf. Chun & Potter, 1995). For example, experiments by Nakayama & Mackeben (1989) and Muller & Rabbitt (1989) studied stimulus driven attention at a particular spatial location and found that it is maximally active approximately 100 ms after stimulus onset and decays thereafter. Numerous computational models have used an attentional function with similar temporal characteristics to explain lag-1 sparing in the attentional blink (Nieuwenhuis et al. 2005; Bowman & Wyble 2007; Shih 2008; Olivers & Meeter 2009, Bowman, Wyble, Chennu, & Craston 2009). A recent series of experiments (Wyble, Bowman & Potter, 2009) has found evidence to suggest that a spatial form of transient attention can be triggered by targets presented among distractors just as those used in RSVP studies. In a manner very similar to contingent capture (Folk, Remington & Johnston 1992), categorically defined RSVP targets cue attention to their own spatial location, enhancing processing of a subsequent T2 at that location if it arrives with a TOA of 100 ms, while impairing processing of a T2 at other locations in the same temporal window. This effect is even present when the T1 is not statistically predictive of the T2's location.

What the results of Wyble, Bowman & Potter (2009) suggest is that exogenous forms of cueing, attentional capture effects, and contingent capture effects may reflect the initiation of attentional episodes at particular spatial locations. Further exploration of this idea awaits experiments that present two or more targets at different locations to determine how attentional episodes bridge or perhaps migrate between different spatial locations in response to the onset of salient or task relevant information.

Episodes at slower presentation speeds

In memory research, it has long been known that there are grouping benefits observed when clustering stimuli together using time (Postle 2003), category (Sharps Wilson-Leff & Price 1995) or spatial proximity (Parmentier, Andres, Elford, Jones 2006). A similar effect is characteristic of the attentional episodes described here, although it is not yet clear how much these various effects have in common. Memory experiments typically present many to-be-remembered stimuli for as long as one second each, ensuring that each one is fully perceived by the subject. The behavioral results tend to produce a U shaped function (i.e. exhibit both primacy and recency). In contrast, during whole report RSVP experiments in which targets are presented at about ten items per second, behavioral performance frequently exhibits an exaggerated primacy pattern compared to slower presentation (Coltheart, Mondy, Dux & Stephenson 2004; Nieuwenstein & Potter 2006). Despite this difference in primacy between RSVP and the much slower form of presentation used in memory experiments, there is commonality in the way that temporal clustering can enhance performance. This suggests that the temporal dynamics of attention, as simulated by eSTST (Wyble et al. 2009a) and other models of temporal attention (Olivers & Meeter 2008; Nieuwenhuis, Gilzenrat, Holmes & Cohen 2005; Shih 2005), may exist across a broader range of time scales than previously assumed.

Conclusion

The eSTST model simulates an episodic form of attention, and experimental evidence supports its predictions both qualitatively and quantitatively. These data illustrate the finding that if task relevant visual stimuli are presented in tight clusters, participants reliably report more of them than when they are interleaved with distractors. This effect is suggested to be due to an attentional mechanism that is best suited to processing stimuli presented in clusters no more than three items in length, and separated by gaps of several hundred milliseconds or more.

In more natural viewing conditions, an analog of these episodes may be continuous sequence of attended visual input: the hand gesture of a magician, the kick of a ball, a passing vehicle, the approach of a person, or the reading of a grammatical unit of text, such as a clause. If this hypothesis is correct, brief periods of attentional suppression occur between episodes, passing unnoticed during natural viewing because most perceptible real world objects persist for periods of at least several hundred milliseconds. However, this suppression produces an attentional blink in a controlled laboratory setting in which targets are masked after abnormally brief durations such as 50ms or 100ms. Under natural viewing conditions, when stimuli are available in the environment for several hundred milliseconds per fixation, this suppression of attention may not result in the loss of much information, yet still provide an important cognitive benefit in punctuating the endpoints of temporal units of visual input.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Akyürek EG, Hommel B. Target integration and the Attentional Blink. Acta Psychologica. 2005; 119:305–314. [PubMed: 15939028]

Arnell KM, Jolicoeur P. The attentional blink across stimulus modalities: Evidence for central processing limitations. Journal of Experimental Psychology: Human Perception and Performance. 1999; 25:630–648.

Bowman H, Wyble B. The Simultaneous Type, Serial Token Model of Temporal Attention and Working Memory. Psychological Review. 2007; 114(1):38–70. [PubMed: 17227181]

Bowman H, Wyble B, Chennu S, Craston P. A reciprocal relationship between bottom-up trace strength and the attentional blink bottleneck: Relating the LC-NE and ST2 models. Brain Research. 2008 April.1202(25–42) 2008.

Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]

Broadbent DE, Broadbent MH. From detection to identification: response to multiple targets in rapid serial visual presentation. Perception and Psychophysics. 1987; 42(2):105–113. [PubMed: 3627930]

Caldwell-Harris CL, Morris AL. Fast Pairs: A visual word recognition paradigm for measuring entrenchment, top-down effects, and subjective phenomenology. Consciousness and Cognition. 2008; 17:1063–1081. [PubMed: 18948037]

Chun MM. Types and tokens in visual processing: a double dissociation between the attentional blink and repetition blindness. Journal of Experimental Psychology: Human Perception and Performance. 1997; 23(3):738–755. [PubMed: 9180042]

Chun MM, Potter MC. A two-stage model for multiple target detection in rapid serial visual presentation. Journal of Experimental Psychology: Human Perception and Performance. 1995; 21(1):109–127. [PubMed: 7707027]

Coltheart V, Mondy S, Dux PE, Stephenson L. Effects of orthographic and phonological word length on memory for lists shown at RSVP and STM rates. Journal of Experimental Psychology: *Learning, Memory & Cognition*. 2004; 30:815–826.

Davelaar EJ, Goshen-Gottstein Y, Ashkenazi A, Haarman HJ, Usher M. The demise of short-term memory revisited: Empirical and computational investigations of recency effects. Psychological Review. 2005; 112:3–42. [PubMed: 15631586]

Dehaene S, Sergent C, Changeux JP. A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proceedings of the National Academy of Sciences, USA. 2003; 100:8520–8525.

Dell Acqua R, Jolicoeur P, Luria R, Pluchino P. Re-evaluating encoding-capacity limitations as a cause of the attentional blink. Journal of Experimental Psychology: Human Perception and Performance. 2009; 35(2):338–351. [PubMed: 19331492]

Di Lollo V, Kawahara J, Ghorashi S, Enns J. The attentional blink: Resource depletion or temporary loss of control? Psychological Research. 2005; 69(3):191–200. [PubMed: 15597184]

Dux PE, Asplund CL, Marois R. An attentional blink for sequentially presented targets: Evidence in favor of resource depletion accounts. Psychonomic Bulletin & Review. 2008; 15:809–813. [PubMed: 18792508]

Dux PE, Marois R. The attentional blink: A review of data and theory. Attention, Perception & Psychophysics. 2009; 71:1683–1700.

Estes WK. Processes of memory loss, recovery, and distortion. Psychological Review. 1997; 104(1): 148–169. [PubMed: 9009883]

Evans K, K, Treisman A. Perception of Objects in Natural Scenes: Is it Really Attention Free? Journal of Experimental Psychology: Human Perception and Performance. 2005; 31(6):1476–1492. [PubMed: 16366803]

Folk CL, Remington RW, Johnston JC. Involuntary covert orienting is contingent on attentional control settings. Journal of Experimental Psychology: Human Perception and Performance. 1992; 18:1030–1044. [PubMed: 1431742]

Garner WR. The accuracy of counting repeated short tones. Journal of Experimental Psychology. 1951; 41:310–316. [PubMed: 14850647]

Haberlandt K, Graesser AC. Processing of new arguments at clause boundaries. Memory & Cognition. 1989; 17(2):186–193.

Hommel B, Akyürek EG. Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. Q. J. Exp. Psychol. A. 2005; 1(1-1)

Just MA, Carpenter PA. A theory of reading: from eye fixations to comprehension. Psychological Review. 1980; 87(4):329–354. [PubMed: 7413885]

Kanwisher NG. Repetition blindness: type recognition without token individuation. Cognition. 1987; 27(2):117–143. [PubMed: 3691023]

Kawahara J, Kumada T, DiLollo V. the attentional blink is governed by a temporary loss of control. Psychonomic Bulletin and Review. 2006; 13(5):886–890. [PubMed: 17328390]

Lamme VAF, Roelfsema PR. The distinct modes of vision offered by feedforward and recurrent processing. Trends in Neurosciences. 2000; 23:571–579. [PubMed: 11074267]

Martens S, Munneke J, Smid H, Johnson A. Quick minds don't blink: Electrophysiological correlates of individual differences in attentional selection. Journal of Cognitive Neuroscience. 2006; 18:1423–1438. [PubMed: 16989545]

Martens, Wyble. The attentional blink: Past, present, and future of a blind spot in perceptual awareness. Neuroscience & Biobehavioral reviews. 2010

Mozer MC. Types and Tokens in Visual Letter Perception. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15(2):287–303. [PubMed: 2525600]

Muller HJ, Rabbitt PM. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15(2):315–330. [PubMed: 2525601]

Nakayama K, Mackeben M. Sustained and transient components of focal visual attention. Vision Research. 1989; 29(11):1631–1647. [PubMed: 2635486]

Nieuwenhuis S, Gilzenrat MS, Holmes BD, Cohen JD. The role of the locus coeruleus in mediating the attentional blink: A neurocomputational theory. Journal of Experimental Psychology: General. 2005; 134(3):291. [PubMed: 16131265]

Nieuwenstein MR, Chun MM, van der Lubbe RH, Hooge IT. Delayed attentional engagement in the attentional blink. Journal of Experimental Psychology: Human Perception and Performance. 2005; 31(6):1463–1475. [PubMed: 16366802]

Nieuwenstein MR, Potter MC. Temporal limits of selection and memory encoding: A comparison of whole versus partial report in rapid serial visual presentation. Psychological Science. 2006; 17(6): 471–475. [PubMed: 16771795]

Nieuwenstein MR, Potter MC, Theeuwes J. Unmasking the attentional blink. Journal of Experimental Psychology: Human Perception and Performance. 2009; 35(1):159–169. [PubMed: 19170478]

Nieuwenstein M, Van der Burg E, Theeuwes J, Wyble B, Potter M. Temporal constraints on conscious vision: On the ubiquitous nature of the attentional blink. Journal of Vision. 2009; 9(18):1–14. [PubMed: 19761351]

Olivers CNL, Meeter M. A Boost and Bounce theory of temporal attention. Psychological Review. 2008; 115:836–863. [PubMed: 18954206]

Olivers CNL, van der Stigchel S, Hulleman J. Spreading the sparing: against a limited-capacity account of the attentional blink. Psychological Research. 2007:1–14. [PubMed: 16317566]

Olivers, Hulleman; Spalek, Kawahara; Di Lollo. The sparing is far from spurious: Reevaluating within-trial contingency effects in the attentional blink. Journal of Experimental Psychology: Human Perception and Performance. (in press).

Ouimet C, Jolicoeur P. Beyond Tak 1 difficulty: The duration of T1 encoding modulates the attention blink. Visual Cognition. 2006

Parmentier FBR, Andes P, Elford G, Jones DM. Organization of visuo-spatial serial memory: interaction of temporal order with spatial and temporal grouping. Psychological Research. 2006; 70:200–217. [PubMed: 15844005]

Pechenkova E. Measuring accommodation of visual attention: Titchener's 'attention-wave' reconsidered? [Abstract]. Journal of Vision. 2006; 6(6):218a.

Postle BR. Context in verbal short-term memory. Memory & Cognition. 2003; 31(8):1198–1207.

Potter MC, Chun MM, Banks BS, Muckenhoupt M. Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1998; 24:979–992.

Potter MC, Nieuwenstein MR, Strohminger N. Whole report versus partial report in RSVP sentences. Journal of Memory and Language. 2008; 58:907–915. [PubMed: 18568099]

Potter MC, Staub A, O Connor DH. The time course of competition for attention: Attention is initially labile. Journal of Experimental Psychology: Human Perception and Performance. 2002; 28(5): 1149–1162. [PubMed: 12421061]

Potter MC, Wyble B, Pandav R, Olejarczyk J. Picture Detection in RSVP: Features or Identity? Journal of Experimental Psychology: Human Perception and Performance. 2010; 36(6):1486–1494. [PubMed: 20695696]

Raymond JE, Shapiro KL, Arnell KM. Temporary suppression of visual processing in an RSVP task: an attentional blink? Journal of Experimental Psychology: Human Perception and Performance. 1992; 18(3):849–860. [PubMed: 1500880]

Rayner K, Kambe G, Duffy SA. The effect of clause wrap-up on eye movements during reading. Quarterly Journal of Experimental Psychology A. 2000; 53(4):1061–1080.

Reeves A, Sperling G. Attention gating in short-term visual memory. Psychological Review. 1986; 93(2):180–206. [PubMed: 3714927]

Sharps MJ, Wilson-Leff CA, Price JL. Relation and item-specific information as determinants of category superiority effects. Journal of General Psychology. 1995; 122(3):271–285. [PubMed: 7650521]

Shih SI. The attention cascade model and attentional blink. Cognitive Psychology. 2008; 56(3):210–236. [PubMed: 17624321]

Sperling G, Weichselgartner E. Episodic Theory of the Dynamics of Spatial Attention. Psychological Review. 1995; 102(3):503–532.

Su L, Bowman H, Barnard P, Wyble B. "Process Algebraic Modelling of Attentional Capture and Human Electrophysiology in Interactive Systems". Formal Aspects of Computing. 2009; 21:513–539.

Taatgen NA, Juvina I, Schipper M, Borst J, Martens S. Too much control can hurt: A threaded cognition model of the attentional blink. Cognitive Psychology. 2009; 59:1–29. [PubMed: 19217086]

Titchener, EB. A textbook of psychology. New York: Macmillan; 1910.

Visser T. T1 Difficulty and the Attentional Blink: Expectancy versus Backward Masking. *Q J Exp Psychol* (Colchester). 2007 Jul; 60(7):936–951. [PubMed: 17616912]

Visser T, Bischof WF, Di Lollo V. Attentional switching in spatial and non-spatial domains: Evidence from the attentional blink. Psychological Bulletin. 1999; 125:458–469.

Vogel EK, Luck SJ. Delayed working memory consolidation during the attentional blink. Psychonomic Bulletin & Review. 2002; 9(4):739–743. [PubMed: 12613677]

Weichselgartner E, Sperling G. Dynamics of automatic and controlled visual attention. Science. 1987; 238(4828):778–780. [PubMed: 3672124]

Wyble B, Bowman H. A neural network account of binding discrete items into working memory using a distributed pool of flexible resources [Abstract]. Journal of Vision. 2006; 6(6):33. 33a.

Wyble B, Bowman H, Nieuwenstein M. The Attentional Blink provides Episodic Distinctiveness: Sparing at a Cost. Journal of Experimental Psychology: Human Perception and Performance. 2009a; 35(2):324–337. [PubMed: 19331491]

Wyble B, Bowman Potter. Categorically Defined Targets Trigger Spatiotemporal Attention. Journal of Experimental Psychology: Human Perception and Performance. 2009b; 35(3):787–807. [PubMed: 19485692]
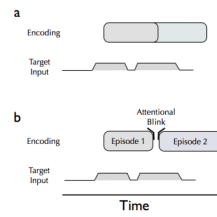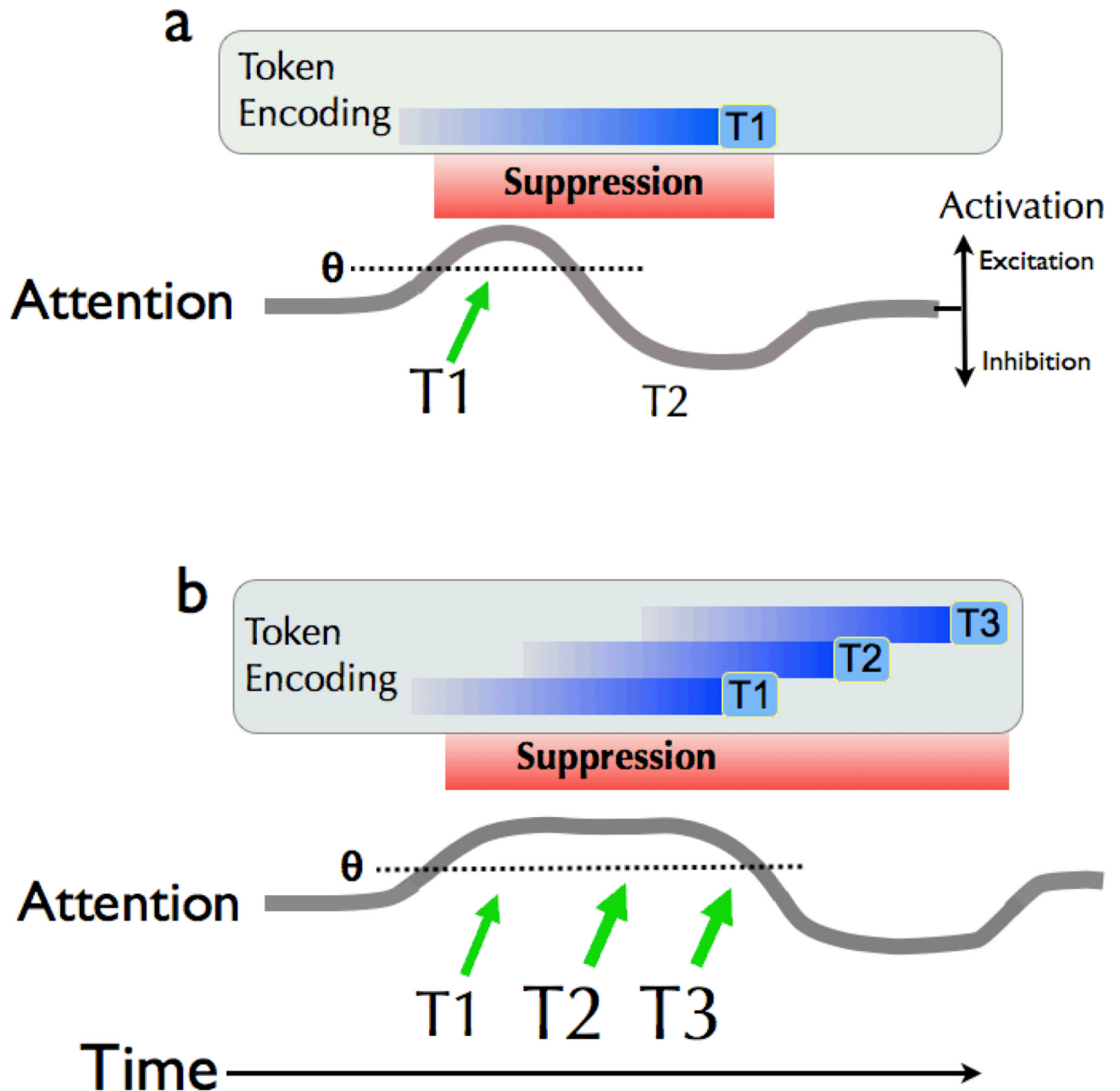
Figure 1.
Two clusters of targets presented closely in time can overlap during encoding (**a**). We suggest that attention is suppressed to provide temporal gaps between periods of encoding to prevent this overlap, creating two distinct attentional episodes (**b**). In this diagram, **Target Input** refers to periods of time during which salient or task relevant visual input is present, and **Encoding** refers to the time course of encoding that information into working memory.
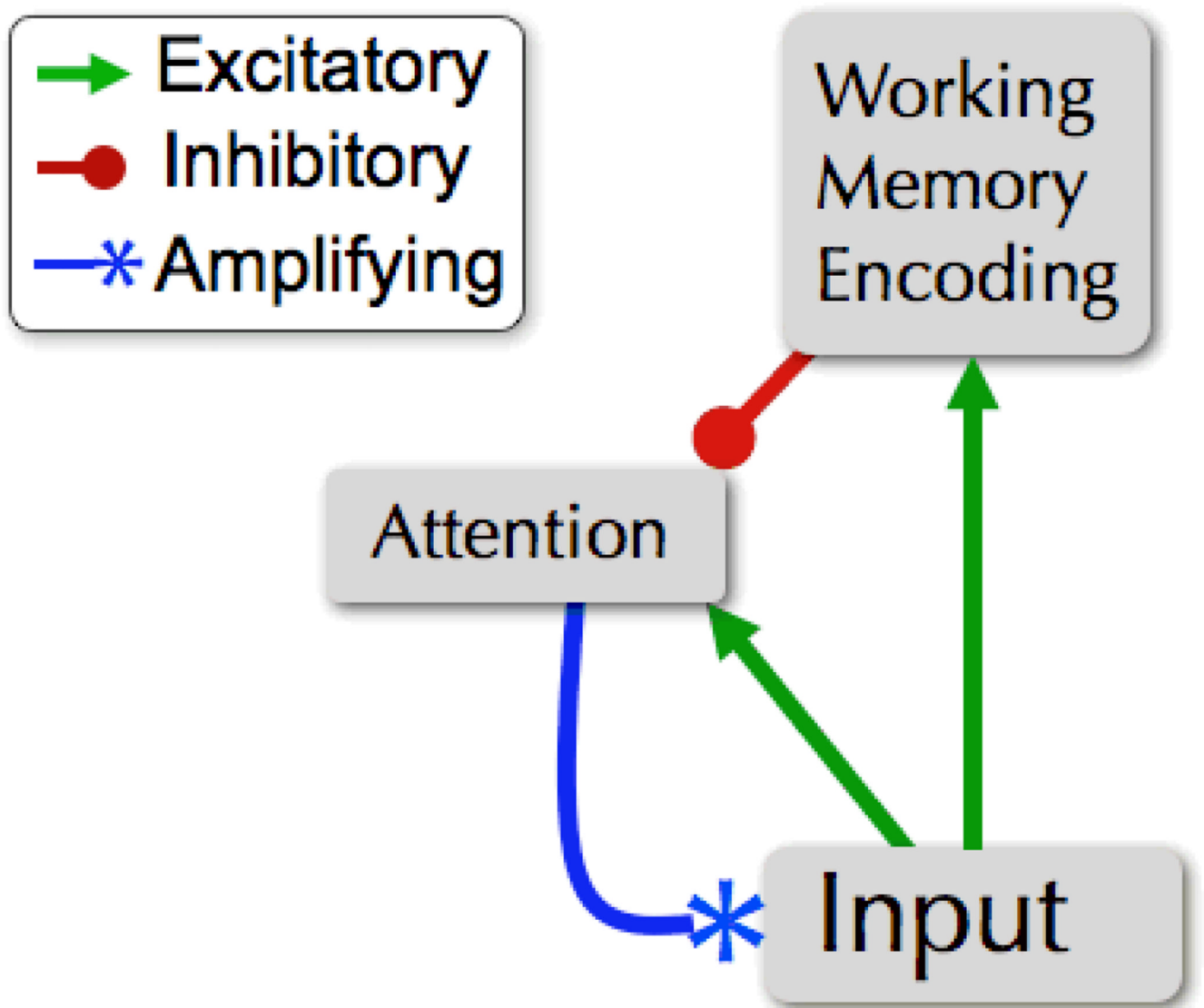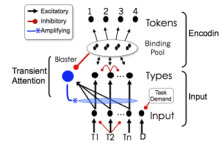
Figure 2.
Competitive regulation of attention in the eSTST model produces episodic attention. In this diagram, attention refers to a stimulus driven process that provides enhancement of visual stimuli at a particular moment of time.

Figure 3.
This figure depicts schematic illustrations of targets being presented to the model (the T's at the bottom of each panel, which are sized in proportion to the amount of attention they receive), and the time course of attention is depicted as the horizontal line. Vertical fluctuations indicate excitation or inhibition of simulated neural activity levels relative to the resting baseline. Activation of attention above the threshold indicated by θ produces enhancement of targets. At the top of each panel is depicted the memory encoding process in which one or more targets can be concurrently stored in memory. During memory encoding, suppressed is applied to the attentional mechanism with different effects depending on the bottom up input. In the top panel, the suppression of attention produces a shut off of

attention between in the interval after T1 producing an attentional blink for T2. In the bottom panel, a chain of successive targets holds the attentional gate open and allows all of them into the encoding stage despite the suppression. Note that suppression of attention begins immediately when the encoding of the first target. However, the bottom up excitation of attention by additional targets can countermand this suppression as long as a recently presented target is active at the input layer.

Figure 4.

The episodic Simultaneous Type/Serial Token (eSTST) model. The model is composed of simulated neural elements divided into two stages: *input* and *encoding*. At the bottom of the input stage, targets and distractors activate corresponding nodes in the attentional layer. The task demand node passively inhibits distractors, but allows targets to trigger attention at a node referred to as the blaster. This attentional deployment enhances the strength of target representations and allows active targets to activate their corresponding type nodes. Active type nodes trigger an encoding process which ultimately results in a tokenized representation of the target being stored in the binding pool. Multiple targets can be encoded in this way, and the order or token allocation corresponds to the perceived order of the targets. Encoding also provides top down suppression of attention, and the competition between the bottom up and top down pressures on attention drives the episodic behavior of the model.
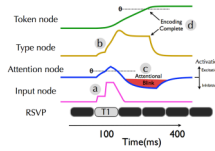
Figure 5.
Simulation of a single trial containing the sequence DTDDDD presented at 100ms per item as shown by the ovals at the bottom of the figure. The four horizontal traces above these ovals depict time aligned traces of the activation level nodes at different levels of processing. In each trace, vertical fluctuations indicate excitation or inhibition of simulated neural activity levels relative to the resting baseline. The token and attention nodes have output thresholds and these are indicated by dotted lines labelled **θ**. These nodes produce no output unless activation exceeds that threshold. The sequence of events is such that a target first triggers attention (**a**), which activates a type node and begins the process of encoding (**b**). Ongoing encoding suppresses attention (**c**) and ultimately results in a stored representation of the target at which point the type node activation falls back to baseline (**d**). Refer to the text for a more detailed description of this encoding process.
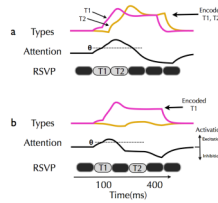
Figure 6.
Time course of the encoding of two sequential targets. Each panel depicts a simulation of the presentation of two targets, along with activation traces of attention and the type nodes corresponding to T1 and T2. The horizontal dotted lines depict the activation threshold **θ** of attention, below which attention is inactive. In panel **a** is depicted a lag 1 trial. Attention triggered by the first target boosts processing of the second target also, such that both targets are encoded. The moment at which encoding is complete is indicated by the arrow in the upper right, and corresponds to the point at which the tokens (not shown) cross threshold. The temporal order in which the activation of the type nodes ends corresponds to the perceived order of the two targets (T1 then T2). In panel **b**, two targets are separated by a 100ms distractor. During this gap, attention is suppressed by T1 encoding so as to defer the processing of T2 into a second episode. However, because the T2 is quickly masked by a following distractor, this deferral of T2 processing produces an attentional blink that swallows the T2. In the absence of a mask, the T2 would be encoded after T1 encoding was complete.
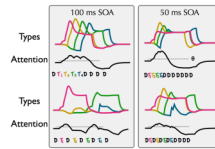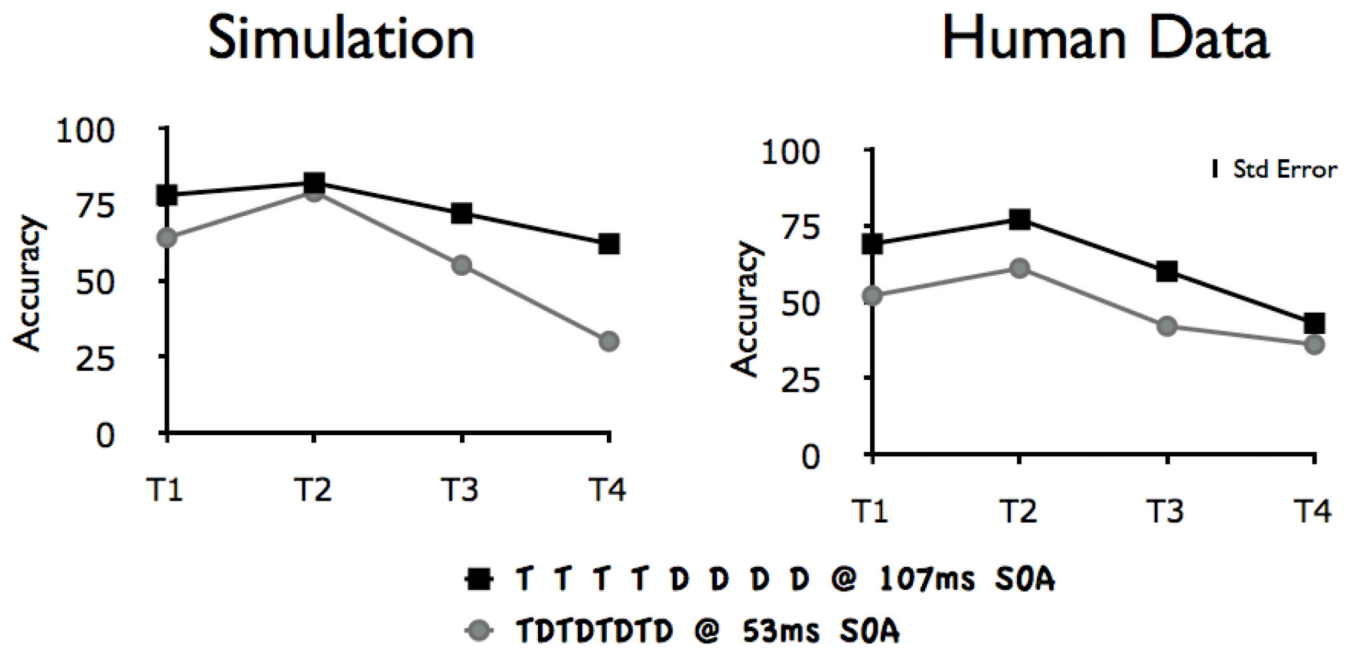
Figure 7.
Simulation of a single trial in each of the four conditions of Experiment 1. Traces reflect activation of 4 type nodes T1,T2,T3 and T4 in addition to the activation level of attention. An attentional episode is active whenever attention is above the threshold **θ**. In the top panels, four consecutive targets are encoded at presentation rates of both 100ms and 50ms per item. In the bottom panels distractors are presented between the targets. In the lower left panel, distractors separate the targets for a sufficient amount of time that two distinct episodes are formed by attention, at the loss of T2 and T4. In contrast, at 50ms per item shown at the bottom right, the distractors are short enough that only one episode is triggered, but the weaker representations of the faster targets results in the loss of T4. Target strength was fixed at 0.9 for each of these simulations, but was varied over a range of values in simulating entire experiments.
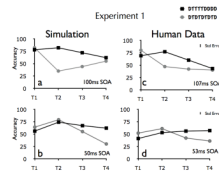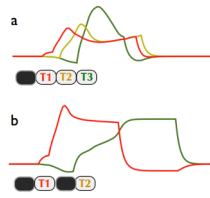
Figure 8.
Comparison of simulated model output and human data in the four conditions of Experiment 1.
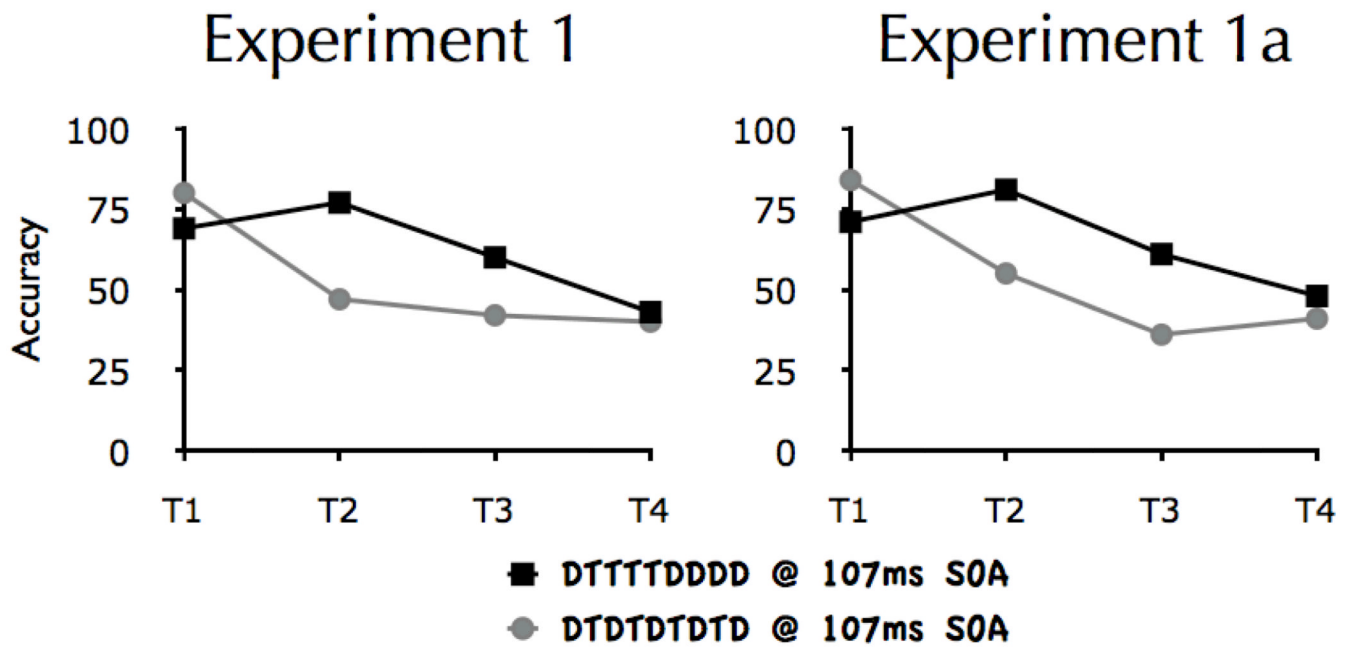
Experiment 1

Simulation

Human Data

Figure 9.
Comparison of simulation and human data between two conditions with equivalent spacing of target onsets. Simulated SOAs were 50ms and 100ms.
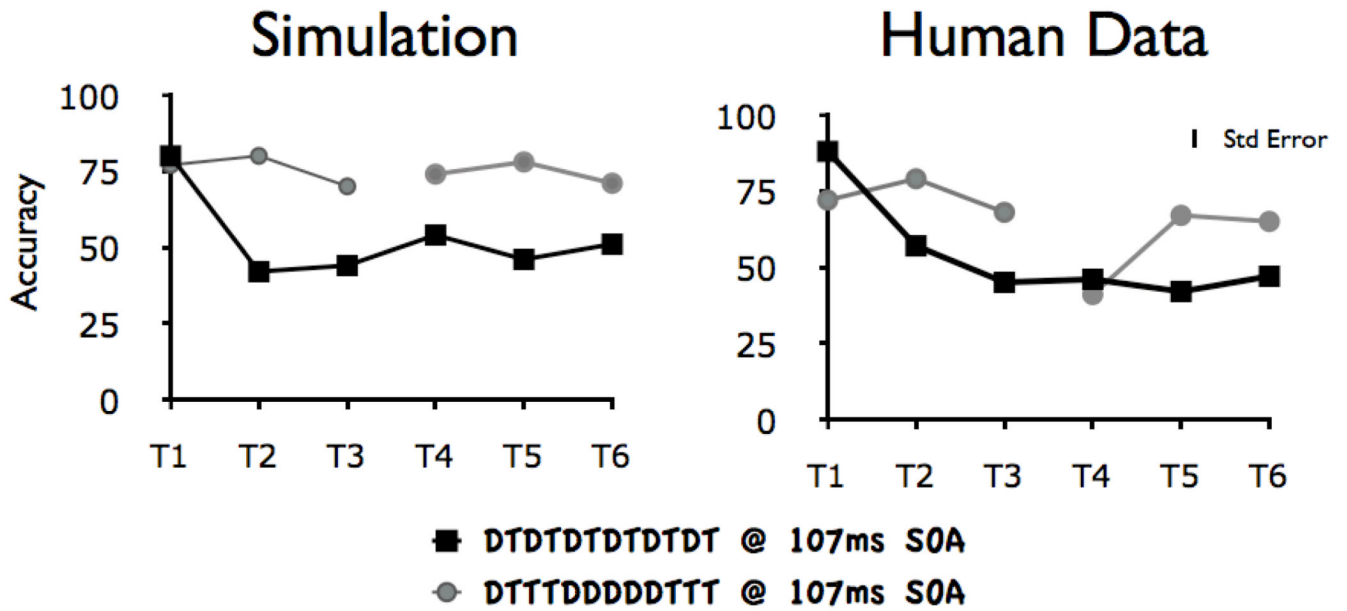
Figure 10.
In simulation, three contiguous targets are encoded as one episode (**a**) and targets presented as noncontiguous are divided into two episodes (**b**). The only difference between these conditions is the presence of an intervening target in panel **a**, which allows attention to be sustained. The cost of combining multiple targets into a single episode can be observed as a temporal order error in panel **a**. T3 has greater strength and completes encoding before the T1 so that the encoded order of the targets is T3,T1, T2. In **b**, T2 has exactly the same strength and relative TOA as the T3 in **a** but attentional suppression forces its encoding to wait until T1 is finished.
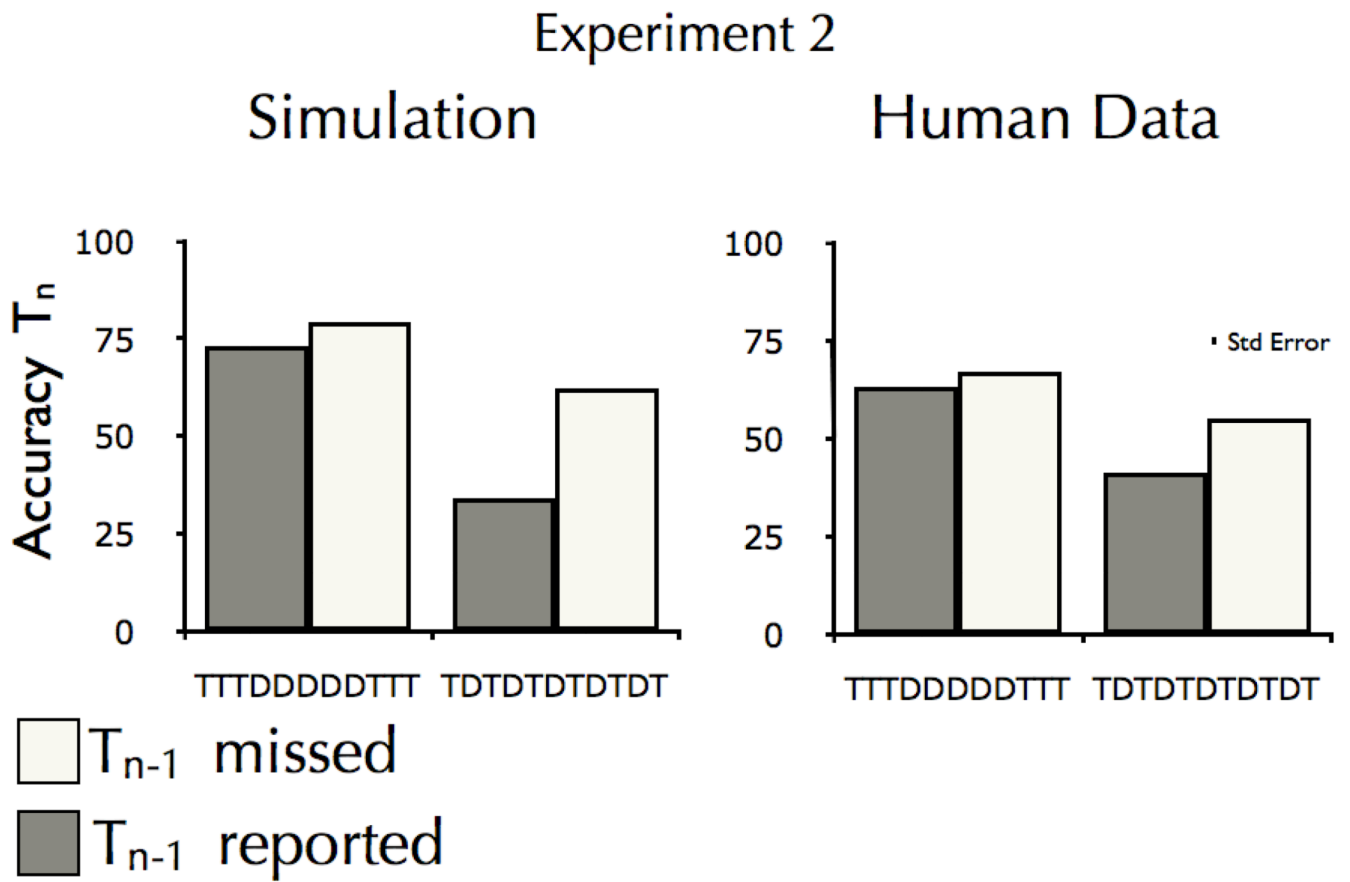
Figure 11.
Comparison of accuracy between equivalent conditions in Experiment 1 and Experiment 1a.

Experiment 2

Simulation

Human Data

■ DTDTDTDTDTDT @ 107ms SOA
**● DTTTDDDDDTTT @ 107ms SOA**

Figure 12.
Comparison of the model with human performance in Experiment 2 for six targets in the two conditions shown in the legend. For the data indicated by grey traces, targets were clustered into two groups, separated by about 500ms.

Experiment 2

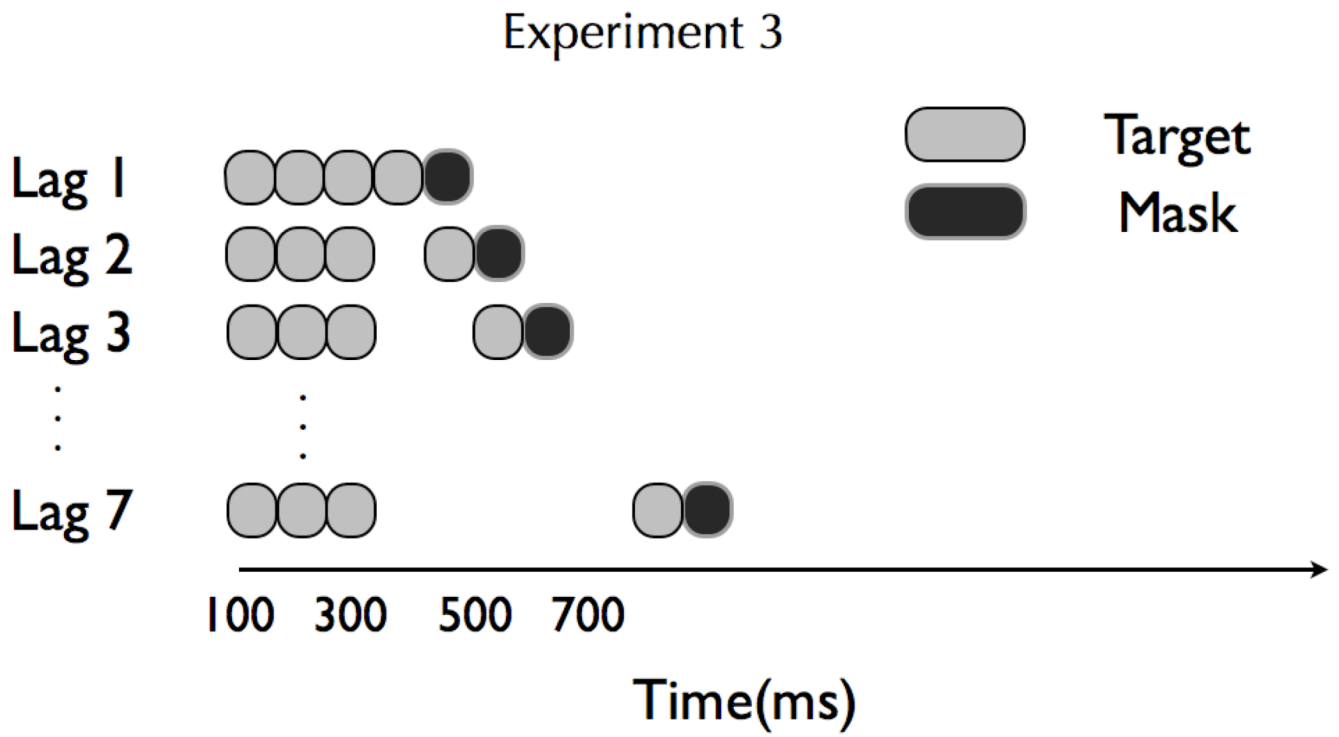## Simulation   ## Human Data

Figure 13.
Encoding of each target as a function of whether the previous target was seen or missed in the two conditions in simulation and in the results of Experiment 2.
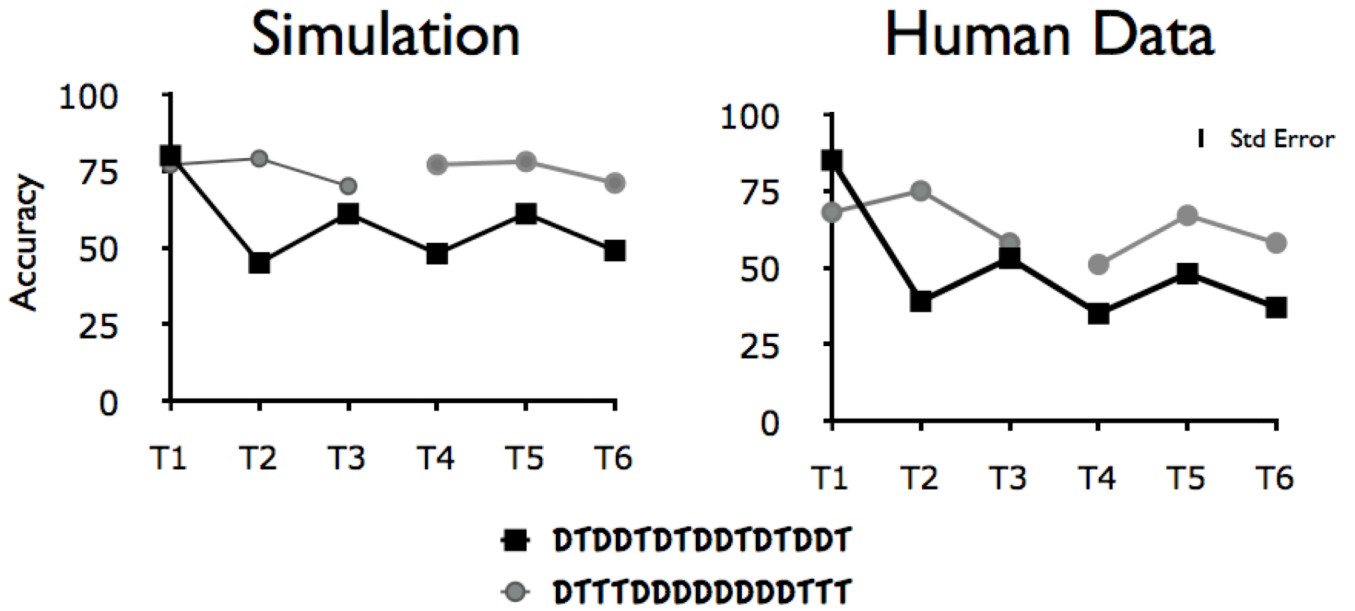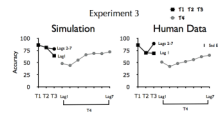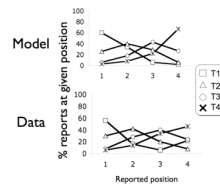
Figure 14.
Experiment 2a replicates the general finding of Experiment 2, but adds three extra distractors between the two sets of targets in the clustered condition.

Figure 15.
The seven conditions used in Experiment 3, varying target clustering and SOA independently.

Figure 16.
Experiment 3 measures the attentional blink created by processing of a 3 target episode without intervening distractors.

Figure 17.
The pattern of temporal positions for both the model and the human data when four targets are presented in a single cluster within an RSVP stream. These graphs illustrate a pattern of migration errors between adjacent items in both the data and the eSTST model that are characteristic of perturbation models such as Estes (1997). This figure is reprinted from Wyble et al (2009).

Table 1

Accuracy scores for targets in conditions from the reported results.

	T1	T2	T3	T4		T5	T6	
TTTT:	69	77	60	43	(stderr = 4 – 6%)			
(Exp 1)								
TTTDDDDTTT:	62	79	68	......		41	67	65	(stderr = 2–3%)
(Exp 2)								
TTTDDDDDDDTTT:	68	75	58	....		51	67	58	(stderr = 2–5%)
(Exp 2a)								