Citation for published version

Bowman, Howard and Wyble, Brad (2005) Computational Modelling of the Attentional Blink. In: Cangelosi, Angelo and Bugmann, Guido and Borisyuk, Roman, eds. Proceedings of the Neural Computation and Psychology Workshop. World Scientific Publication ISBN 981-256-324-5.

DOI

Link to record in KAR

https://kar.kent.ac.uk/14375/

Document Version

UNSPECIFIED

MODELLING THE ATTENTIONAL BLINK

HOWARD BOWMAN[A] BRADLEY WYBLE[B]

[A] *Computing Laboratory, University of Canterbury*
Canterbury, Kent CT2 7NF, UK
hb5@kent.ac.uk

[B] *Computing Laboratory, University of Canterbury*
Canterbury, Kent CT2 7NF, UK
bw5@kent.ac.uk

This paper describes a computational model of the Attentional Blink constructed using the dual-stage model proposed by [1] and also incorporating a token based account of working memory [2]. This model reproduces data from a number of blink paradigms and makes predictions that lag-1 sparing is temporal and not sequential in origin. A further prediction is that enhancing the distinctiveness of T2 can impair T1 performance and also provoke order inversions of T1 and T2. Experiments from our lab examined the validity of these predictions. Implications and results are discussed.

1. Introduction

Many studies of temporal attention have employed Rapid Serial Visual Presentation (RSVP), in which stimuli are presented one after the other at the same spatial location. At presentation rates of 10 items per second, targets in the stream are frequently missed. The Attentional Blink (AB) [3,1] paradigm uses such an RSVP stream, with two targets (denoted T1 and T2). The basic finding, depicted in figure 2b, is that there is a period of approximately 500 msec during which processing of T1 seems to impair the ability to detect and report T2. This suggests that the deployment of attention to processing T1 has a temporal window of a little over half a second. This interpretation is complicated by lag 1 sparing, which is the robust finding of almost unimpaired performance on T2 when it immediately follows T1. Lag 1 sparing raises the central question of why attentional resources are too limited to process both T1 and T2 at lags of 100-600 ms, but are sufficient if T2 is at shorter intervals, providing a clear challenge to computational modelling of the AB.

This paper responds to these issues by proposing a computationally detailed model that reproduces a spectrum of experimental data on what can be argued to be a canonical AB task, in which the subject must report the two letter targets that appear in a stream of digit distractors [1]. Specifically, in the context of this task, our model reproduces a core set of findings, that is (1) a basic blink effect with lag 1 sparing; (2) an impairment in T1 performance at lag 1; (3) an increase in temporal order confusion at lag 1, and to a lesser extent at lag 2; and (4) effects of unmasking T1 and T2, i.e. attenuation of the blink if one or more blanks are inserted after T1 and / or T2 in the RSVP stream.

Two key principles underlie the model. Firstly, in accordance with [4] and [5], we subscribe to the position that a major source of competition between RSVP items is at the level of visual features. Thus, items in the stream are backward masked by the items that immediately follow them. This position is supported by the attenuation of the blink when targets are unmasked, i.e. are followed by blanks. In our model, masking determines the strength of the activation trace of a target. Thus, weak traces arise from masked targets (e.g. as occur in the basic blink condition), while strong traces arise when targets are unmasked (e.g. when they are followed by blanks). Critically, in our model, “bottom-up” activation strength generated by a stimulus determines how easily the representation of that stimulus can be consolidated into working memory and also, the length of time that attentional resources are occupied with this process. As a result, the severity of the blink bottleneck is regulated by bottom-up trace strength, which is in turn determined by the level of masking to which a target is subjected.

The second key principle that underlies our model is the types – tokens distinction [2]. Types in our model are detailed representations of the identity of an item, including semantic and perceptual features. However, types are impoverished in their representation of how and when an item occurred. Consequently, it is also necessary to represent token information, which records instance specific details of the occurrence of an item. In particular, in our model, tokens are compact working memory encodings, which record both how and when the item occurred (e.g. its temporal position relative to other items) and which enable type information to be regenerated during retrieval. Our implementation of this model encapsulates this token binding system into a dual-stage model as described by [1,6,7].

An earlier version of this model was presented in the proceedings of the 2003 NCPW conference [8]. This model improves on that work through the implementation of a transient attentional resource that creates a temporal window of enhanced salience, which allows it to address other types of data. Also, we now separate semantic and task-related processing into different layers with different time courses of activation, an essential step in allowing targets and distractors to prime one another.

We begin this paper by outlining the technical details that underlie our model in section 2. Then in section 3 we compare human data with the results of our simulations. Finally, in section 4 we discuss the implications of our work, and compare the model's predictions with experimental work in our lab.

2. Methods

2.1. *Overview*

This model implements a representation of the processing of an RSVP stream containing 16 distractors and 2 targets presented at a simulated rate of 100 msec per item (Stimulus Onset Asynchrony: SOA). Targets are separated by 0-6 distractors. The model simulates the neural dynamics associated with the encoding of the targets into tokens that support later retrieval.

2.2. *Neural Elements*

This model uses neural elements designed to broadly capture the generalized excitatory and inhibitory dynamics of cortical areas. Activation functions are a combination of bias, excitation, inhibition and leak currents. Connections between elements are excitatory or inhibitory and are not modifiable, with the exception of binding links described below. Outputs from each neural element are derived from thresholds combined with a sigmoid-like function. There are no synaptic delays. The activation function for the membrane potential of all neural elements is show in equation 1,

$$\begin{aligned}MP_{(i,j,t)} &= MP_{(i,j,t-1)} + DT\_VM_{(i)} * ((Bias_{(i)} + Excite_{(i,j,t-1)}) * (EE_{(i)} - MP_{(i,j,t-1)}) + \\ &Inhib_{(i,j,t-1)} * (EI_{(i)} - MP_{(i,j,t-1)}) + Leak_{(i)} * (EL_{(i)} - MP_{(i,j,t-1)}))\end{aligned} \tag{1}$$

where $MP_{(i,j,t)}$ is the membrane potential for neuron *i* of layer *j* at time *t*, affected by *Bias*, *Excite*, *Inhib* and *Leak* with a time constant *DT_VM*. *EE*, *EI* and *EL* represent reversal potentials. This membrane potential is used in the following output function which simulates a sigmoid function that bounds a neuron's output to the range [0,1), where θ and γ represent threshold and scaling parameters.

$$Out_{(i,j,t)} = \frac{[MP_{(i,j,t)} - \theta_{(j)}]_+ * \gamma_{(j)}}{[MP_{(i,j,t)} - \theta_{(j)}]_+ * \gamma_{(j)} + 1} \quad \text{where } [x]_+ = \begin{matrix} x & if \quad x > 0 \\ 0 & otherwise \end{matrix} \tag{2}$$

Apart from the recall and binding link mechanisms, described in more detail below, all elements of this model are created entirely through a network of these elements. An exhaustive description of the parameter values is beyond the scope of this paper, but will be available online at,

http://www.cs.kent.ac.uk/projects/cncs/online/bw5/ncpw2004/blink/.

This paper focuses on describing and analyzing the functional characteristics of this architecture, which can be obtained over a broad range of parameter values.

2.3. *Structure*

The structure of this model can be viewed as a two-stage architecture similar to that described in [1] in which the first stage is capable of operating in parallel, while the second is serial in nature. The second stage implements a tokenization process that binds tokens to types in a similar manner to that described in [2].

The first stage comprises the bottom 4 layers depicted in Figure 1. Lateral and feedforward inhibition in the first two layers creates a masking effect that reduces the duration and amplitude of masked relative to unmasked items. Traces also differ systematically in strength based on variance applied at the input layer, to simulate varying degrees of featural masking by digit/letter combinations. Representations are localist, with one neuron per item in each layer of the first stage. The representations in this masking layer are passed directly to both of the next two layers (3 and 4), which are intended to represent items at semantic/categorical levels. The first of these layers maintains activation traces for several hundred milliseconds, intended to represent extended semantic processing that is immune to conditions of task demand.

The second of these layers (the Task Selection Layer: TSL) invokes a representation at which cognitive control can select a specific category of items for further processing. Task demand input, representative of such strategic cognitive control, selectively excites neurons corresponding to targets while suppressing neurons corresponding to distractors. The TSL neurons are weakly self-excitatory, but share weak mutually inhibitory connections, providing a measure of interference between co-active representations. A further means by which TSL representations interfere with one another is provided indirectly through the unavailability of an excitatory recurrent mechanism that provides a 50 msec pulse of excitation to all items in the TSL and semantic layers when any neuron in the TSL is activated. The contribution of this enhancement is critical in strengthening weak traces to a level at which they can activate a token and be encoded. Justification for the inclusion of this excitatory recurrent pulse is provided in the discussion.

This excitatory mechanism is strongly suppressed during a token binding process (as described below). These two methods of interference (lateral inhibition in the TSL and suppression of the excitatory pulse) make it difficult for weak items to become activated in the TSL while an existing representation is active.

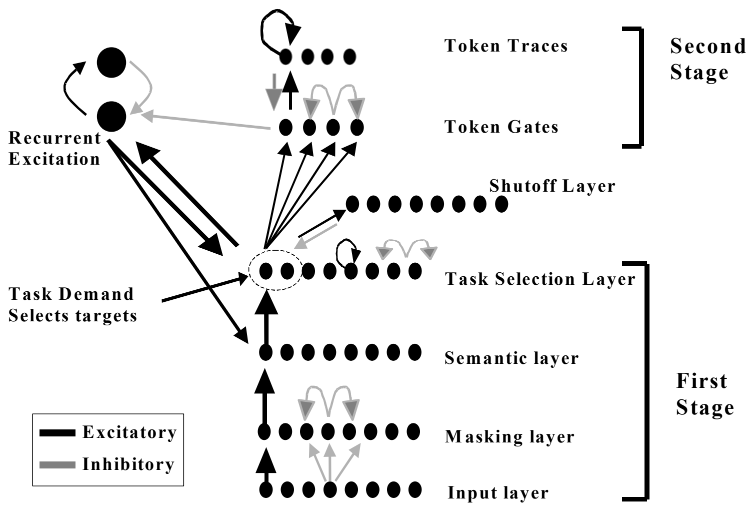

Figure 1. The structure of the model, note that connection patterns between layers are only depicted for some neurons, but apply for the entire layer.

The TSL is also connected in a 1 to 1 fashion with a shutoff layer that contains neurons capable of selectively inhibiting each task selection neuron after a sufficient amount of activity. This mechanism is critical in preventing strong traces with long durations from erroneously encoding multiple tokens, but can also cause repetition blindness for repeated elements [2]. The second stage of this model implements a process of binding working memory tokens to the item "types" in the TSL. During the presentation of the stream, a token may be activated by items in the TSL, which are above threshold. The activation of a token initiates the construction of binding links between that token and any items that are currently active. The processes of activating the token and building these links take significant time (hundreds of msec), and occur faster with stronger traces than weaker ones. Tokenization is implemented through a dual-layer system consisting of gate and trace neurons, arranged in a series of pairs. Together one gate neuron and one trace neuron comprise a token. The gate neurons receive a small bias and are mutually inhibitory such that only one of them is active at any time, ensuring that only one token is available. The trace neurons are strongly self excitatory and thus are designed to switch on once given sufficient input, after which time they self-sustain and inhibit their respective gate neuron, making the bound token unavailable for further binding, and releasing the remaining tokens from inhibition.

Tokens in this model have four states, (1) available, (2) unavailable, (3) binding and (4) bound. An available token (1) has a minimally active gate neuron and an inactive trace neuron. While one token is available, others are unavailable (2) due to lateral inhibition between the gates. A token in the process of binding (3) has a highly active gate neuron, being driven from one or more items in the first stage, and a trace neuron with a steadily accruing level of activation. A token that has already been bound (4) has an inactive gate neuron receiving continuous inhibition from its active trace neuron. While a token is being bound, binding links are incrementally created from gate neurons to individual items in the TSL. The rate of increase of these links is proportional to the strength of the individual traces in the TSL, when the gate is sufficiently active. The sustained activation of the trace neurons are used later to determine which tokens were bound. Binding links from the gate neurons to the TSL determine what item or items were encoded by those tokens. These binding links have no effect on the dynamics of the model during presentation and encoding of the RSVP stream.

This token system has a number of important properties. First, multiple tokens may point to the same type. This feature allows the system to represent multiple instances of the same item. Although not a factor in AB studies that forbid repetition of targets, this feature provides an important facility to a working memory implementation. Second, a single token may point to multiple items. This is a case that may occur when items are presented in close temporal proximity (< 150 msec). Third, the tokens are made available sequentially during encoding, which implicitly encodes order information. Gate neurons compete to become available at the beginning of a trial and also when the current token has been bound. The winner is determined by an ordered pattern of bias currents applied to the gate neurons, which effectively determines which token is first, which is second and so on. In this way, the model preserves the order of item encoding when items are presented slowly enough that a second token is available to encode the second item (ie at 300 msec lag). Fourth, there are a limited number of tokens available, in the case of this model, 2 are sufficient, but 3-4 may exist.

At the conclusion of an RSVP stream, the model simulates a recall phase to disambiguate the erroneous bindings caused by a presentation rate that is too rapid for the token system to accommodate. This simulation is not performed explicitly within the network, as it is only our intent to model encoding processes. For successful recall of a target at the end of a trial, a binding link of sufficient strength must exist from any token to that target, and that token must have an active trace neuron If both targets were successfully bound to either one or two tokens, the recall phase probabilistically determines in which order they were perceived by comparing the relative strength of binding links from each token to each of the two targets. In determining order, we are assuming that a retrieval process makes reasonably efficient use of all available information, including bindings that are below threshold, and tokens with inactive trace neurons. The sum of the token binding links indicating that T1 and T2 are in the correct, or inverted order are compared to determine a probability that an inversion occurred on a given trial. This probability is then compared with a randomly generated number in the range [0,1] to determine whether an inversion occurred on a given trial. This approach simulates a robust disambiguation process that makes sense of “sloppy” token bindings.

3. Results

Figure 2 displays model data alongside human data. This model is able to successfully reproduce all of the qualitative features of the standard blink, T1 + 1 Blank, and T2 at the end of the stream conditions, and also creates some predictions for data that have not yet been collected.

The mechanism by which the model generates a blink is largely a result of suppression of the recurrent excitatory enhancement in layer 4 that occurs during tokenization, as well as lateral inhibition within the TSL. The combination of these effects prevents a T2 from reaching a level of activation sufficient for tokenization. The binding of T1 can take long enough that a weak T2 trace in layer 3 has decayed before it can be bound into a second token.

Lag-1 sparing results from the T2 being close enough in time to T1 to take advantage of the excitatory pulse, allowing it to be encoded alongside T1 into the first token. Via lateral inhibition this dual-encoding of T2 and T1 comes at the partial expense of the T1 binding strength, which is evident in the loss of T1 performance at lag 1. Furthermore, that T1 and T2 are bound to the same token, with only partial information spilling over to a second token, causes a massive increase in order inversions at lag 1 (Figure 2c).

Blanks in the T1+1 slot attenuate the blink by producing a much stronger T1 trace. T1 is then bound rapidly. Consequently, it is easier for strong T2’s to persist in the semantic layer long enough to be bound. When the T2 is unmasked by being the last item it has a particularly strong T2 trace, which can break through or outlast the window of inhibition caused by the T1 token binding.

4. Discussion

This AB model succeeds in causing the emergence of a blink from the dynamics of a neural network that embraces a number of pre-existing theories [1,6]. Further, the contour of this blink is manipulated by modifications of the RSVP stream in a way that is very similar to that shown in human data and makes predictions for what might be observed in variations of the AB paradigm.

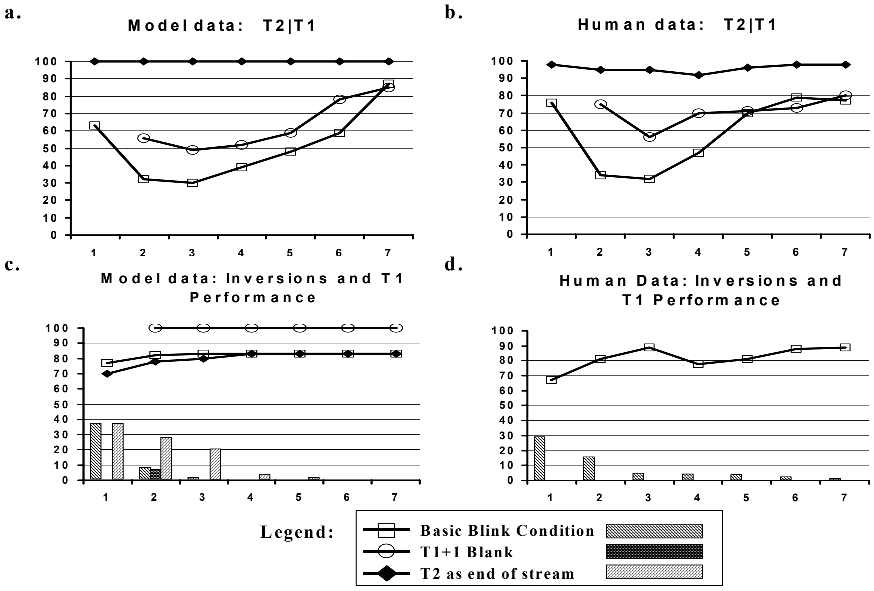

Figure 2a-d: These 4 graphs demonstrate the model's performance (a, c) compared to human data (b, d). The T2 performance data (a, b) represents the accuracy in reporting T2 on trials in which T1 was reported. In c and d, T1 accuracy is shown by the lines at the top of the graph, while the bars at the bottom indicate the percent chance for the reported order of T1 and T2 to be inverted. Human data are from [1] except the T2 end of stream data which is from [9]. Horizontal axis represents the number of intervening distractors, while the vertical is accuracy. The simulated results in panel c include data not yet found in the literature.

One important aspect of this model is the way in which the recurrent excitatory pulse affects all items. This implementation was chosen because it fits available data by [10] demonstrating that a distractor in the T1+1 slot has an ability to prime a T2, while a distractor in the T1+2 slot does not. Our implementation of an excitatory pulse applied to all items achieves this increase in salience restricted to a time slot immediately following a task-relevant item.

In a simulation to test the influence of primes in T1+1 and T1+2 slots, the former was capable of strongly priming T2 while the latter had little effect (63% vs 37% accuracy at lag 3).

The real heart of this model is in the dynamics of token binding, which presupposes that ongoing binding must be protected from interference by ensuing items. This binding process is considered to be promiscuous, in that a token will bind indiscriminately to any active traces within the TSL. If multiple tokens could bind simultaneously and without interference, there would be no need for a protective mechanism such as the one implemented in this model, but the design of such a system would be considerably more complex.

This theoretical position has the further implication that many of the blink effects are temporal in origin, and not the result of the sequential adjacency of T1 and T1+1 slots. Multiple traces can coexist in the TSL, if presented rapidly enough to take advantage of the initial excitatory pulse, and it is their mutual interference that causes order inversions and T1 impairment at lag 1. Any T2 that follows T1 within the appropriate interval (approx 100-150 msec) can join into the binding process of the tokenization initiated by T1. Consequently our model makes the further predictions that at SOA's of 50 msec, items presented at lag 2 should exhibit both sparing as well as order inversions. Preliminary data from our lab for RSVP rates of 10 and 20 items/sec (100 and 50 msec SOA) demonstrates that this is correct. A comparison of the fast and slow curves reveal that the blink curve is a function of time, and not the sequential nature of the presentation. Strong lag-2 sparing was obtained for the 50msec SOA and the maximal depth of the blink was at lag 6.

Another major prediction of this model is that enhancing the strength of T2 increases interference with T1 at lag 1. Through analysis of our existing data we have confirmed this prediction for different classes of letters grouped by featural difficulty. A letter-by-letter analysis of single target RSVP streams reveals that some have consistently higher accuracy rates, presumably because of features that are more easy to recognize through the digit masks. Easy letters were: A H N Y T and L. Difficult letters were B C E P J and V. Other letters with intermediate difficulties were ignored for this analysis.

In an AB experiment, 14 subjects saw random pairings of letters, at lags 1-8, with 92 msec SOA's. Examination of all 8 lags discovered that in the comparison of trials which used easy or hard letters for T2, the easy T2's selectively impaired T1 performance only at lag 1, (lag-1 two-tailed T-test, $p < .006$, ns at other lags). The fact that T2's with presumably strong feature traces interfere with T1 more than T2's with weak traces, and only at lag-1, constitutes empirical validation of our prediction that lag-1 sparing is the result of a single combined token being created for targets presented at near 100 msec SOA.

5. Conclusion

This model succeeds in integrating a token-based theory of working memory with a dual stage model of the AB, and in doing so exposes important issues that need to be explored to further our understanding of this phenomenon. It is our position that much can be learned about the nature of target binding by focusing experiments on the 150 msec following the T1 with SOA's of 50 msec, and also to explore further the effect of T2 manipulations on T1 performance. Specifically, in this theoretical construct and more generally that of the two-stage model, the questions most outstanding regarding the interaction of T1 and T2 are the following: Can multiple tokens be bound simultaneously or is the process serial in nature? Do T1+1 blanks, as well as other means of attenuating the blink cause their effect by reducing the time to bind a token? Do T2+1 blanks attenuate the blink in the same manner? Are T1 and T2 bound together into a conglomerated token at Lag 1? Experiments designed to address these issues will not only improve our understanding of the Attentional Blink phenomenon, but will also shed light on the temporal resolution of visual attention for fleeting objects.

6. References

[1] Chun, M. M. and M. C. Potter (1995). "A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation." J Exp Psychol: HPP **21**(1): 109-127.

[2] Kanwisher, N. G. (1987) "Repetition blindness: Type recognition without token individuation" Cognition, 27:117-143.

[3] Raymond, J.E., K.L. Shapiro et al (1992) "Temporary Suppression of Visual Processing in an RSVP Task: An Attentional Blink" J Exp Psychol, HPP, 18(3):849-860.

[4] Seiffert, A.E., Di Lollo, V. (1997) "Low-Level masking in the Attentional Blink" J Exp Psychol, HPP, 23: 1061-1073.

[5] Maki, W.S., G. Bussard, et al (2003) "Sources of Interference in the Attentional Blink: Target-distractor Similarity Revisited" Perception & Psychophysics, 65(2), 188-201.

[6] Chun, M.M. (1997) "Types and tokens in visual processing: a double dissociation between the attentional blink and repetition blindness" J Exp Psychol, HPP, 23: 738-755.

[7] Shapiro, K. L., J. Driver, et al. (1997). "Priming from the attentional blink: A failure to extract visual tokens but not visual types." Psychological Science ,**8**: 95-100.

[8] H. Bowman, B. Wyble, et al (2004) "Towards a Neural Network Model of the Attentional Blink." Eighth Neural Computation and Psychology, vol. 15 of Progress in Neural Processing, H. Bowman and C. Labiouse (eds), World Scientific, 178-187.

[9] Giesbrecht, B. and Di Lollo, V., (1998) "Beyond the Attentional Blink: Visual Masking by Object Substitution." J Exp Psychol, HPP, 24: 1454-1466.

[10] Chua, F.K., J. Goh et al (2001) "Nature of Codes Extracted During the Attentional Blink" J Exp Psychol, HPP, 27(5), 1229-1242.x