

The Simultaneous Type, Serial Token Model of Temporal Attention and Working Memory

Howard Bowman and Brad Wyble
University of Kent

A detailed description of the simultaneous type, serial token (ST²) model is presented. ST² is a model of temporal attention and working memory that encapsulates 5 principles: (a) M. M. Chun and M. C. Potter's (1995) 2-stage model, (b) a Stage 1 salience filter, (c) N. G. Kanwisher's (1987, 1991) types–tokens distinction, (d) a transient attentional enhancement, and (e) a mechanism for associating types with tokens called the binding pool. The authors instantiate this theoretical position in a connectionist implementation, called neural-ST², which they illustrate by modeling temporal attention results focused on the attentional blink (AB). They demonstrate that the ST² model explains a spectrum of AB findings. Furthermore, they highlight a number of new temporal attention predictions arising from the ST² theory, which are tested in a series of behavioral experiments. Finally, the authors review major AB models and theories and compare them with ST².

Keywords: temporal attention, working memory, attentional blink, neural networks

Supplemental Material: <http://dx.doi.org/10.1037/0033-295X.114.1.38.supp>

INTRODUCTION

1. Temporal Attention

1.1. Background

The study of *attention* (Driver, 2001; Logan, 2004) is focused on understanding how the brain identifies salient items among a set of competing stimuli. The majority of visual attention research has considered how selection occurs when the competing stimuli satisfy two criteria: (a) They are spatially offset from one another, and (b) they are presented simultaneously (Duncan, 1981; Eriksen & Eriksen, 1974; Treisman, 1998). However, attentional processing still obtains even if the above criteria do not apply. Indeed, the problem of how salient items are selected when the competing stimuli are offset in time, rather than space, must also be solved by a cognitive system. However, although there is increasing investigation of *temporal attention*, it remains, especially from theoretical and computational standpoints, considerably less explored than spatial attention.

When temporal attention is explored, it is natural to present stimuli briefly (typically for around 100 ms) with each being replaced by the next at the same spatial location, generating what is called rapid serial visual presentation (RSVP; Chun & Potter, 1995; Kanwisher, 1987; Raymond, Shapiro, & Arnell, 1992; Shih & Sperling, 2002). At this rate of presentation, performance is below ceiling, permitting investigation of factors affecting perception of targets. Furthermore, because all items are presented at the same spatial location, each item (apart from the first) masks the previous item (Giesbrecht, Bischof, & Kingstone, 2003; Giesbrecht & Di Lollo, 1998). Thus, a key characteristic of explorations of temporal attention is that representations of stimuli are fleeting (i.e., only briefly available to the retina).

In addition, although the question of how the brain detects a single salient item in an RSVP stream is interesting, to reveal the characteristics of temporal attention, we need to explore streams containing multiple salient items. This allows a number of key questions to be investigated. For example, how long is attention allocated to an initial target before it is free to be allocated to a second; can a salient item interrupt processing of an earlier item and cause attention to be redirected; and what constitutes salient in this context: relevance to the task at hand, emotional significance, similarity to a target category, and so on?

The brain not only enables selective attention to salient items but also enables higher level cognition (e.g., executive processes, reasoning, language comprehension) to act upon attended items. The basis of such processing is the capacity to encode attended items into working memory (WM). Indeed, when considering temporal attention, the issue of WM encoding is particularly pivotal because not only does this process involve storage of items, it also supports encoding of information about the temporal context in which an item occurred. With RSVP streams, assigning such *episodic contexts* is particularly difficult, because it requires the stream to be “individuated” through time; that is, a discrete episode

The United Kingdom Engineering and Physical Sciences Research Council supported this research (Grant GR/S15075/01). We thank Phil Barnard, Mary Potter, Patrick Craston, Mark Nieuwenstein, Dinkar Sharma, Marvin Chun, David Huber, Kyle Cave, and Randall O'Reilly for discussions and input that have contributed to the development of the ST² model.

The MATLAB code that implements the neural-ST² model is available on the following page: <http://www.cs.kent.ac.uk/people/staff/hb5/ST2model.zip>.

Correspondence concerning this article should be addressed to Howard Bowman or Brad Wyble, Center for Cognitive Neuroscience and Cognitive Systems, University of Kent, Canterbury, Kent CT2 7NF, United Kingdom. E-mail: H.Bowman@kent.ac.uk or B.Wyble@kent.ac.uk

needs to be associated with the occurrence of each item that is encoded into WM.

In support of this position, it is clear that humans can report episodic information about when and how events occurred. In particular, they can report the order in which events arose. Furthermore, temporal order is a key information-bearing dimension, the extraction and recording of which is critical to high-level cognition. Linguistic abilities make this clear; for example, word order in the sequence *John hit Jane* is central to attribution of meaning. Indeed, one of our main theses in this article is that the human cognitive system is good at filtering nonsalient items during visual processing and thus the main constraint on the capacity to direct attention through time is the cost incurred in associating episodic contexts while encoding items into WM.

The basic problem this article focuses on is, given a sequence of fleeting representations of stimuli, how is it that the brain identifies the stimuli that are salient, sustains them, and encodes them into WM. Central to this process is the allocation of discrete episodic contexts during encoding. Specifically, we devise a comprehensive theoretical proposal for this process, which is encapsulated as the simultaneous type, serial token (STST or ST²) model. This theory is realized in a connectionist model, neural-ST², that is constrained both by behavioral data on RSVP tasks (particularly the attentional blink [AB]; Chun & Potter, 1995; Raymond et al., 1992) and by neurophysiology. In order to frame our theory, we identify a set of functional requirements.

1.2. Functional Requirements for the Model

When we consider behavioral data, it becomes clear that in certain very demanding visual environments, some of the following functional requirements will not obtain.¹ Thus, our claim is that the following capacities characterize the normal functioning of temporal attention.

1. *Salience detection.* The system should have the capacity to rapidly detect that a stimulus is salient when a fleeting representation of it arises among a set of (temporally) competing stimuli.
2. *Sustain representations.* The system should have the capacity to sustain fleeting, but salient, representations of visual stimuli, in order that they can be acted on (and encoded into WM).
3. *Ascribe episodic context.* The system should have the capacity to ascribe (discrete) episodic contexts to items as they are encoded into WM, thereby enabling order information to be retrieved.
4. *WM maintenance.* The system should have the capacity to maintain multiple items in WM.
5. *Repetitions.* The system should have the capacity to represent multiple instances of the same item. Although this may seem like a trivial constraint, in fact, it is not always that easily obtained; see Section 6.1. The reason for this difficulty is that perceiving and encoding repetitions requires the neural resources used for detecting the

first instance of an item to be “freed up” before the second instance arises.

The article is organized as follows. In Section 2, we present an informal description of the ST² model, motivating the approach from the perspective of our functional requirements. This is followed, in Section 3, by a discussion of the key empirical findings we model; these are based on the AB phenomenon. Then, in Section 4, we present a formal, neurally explicit, description of the model, called neural-ST², and show how it reproduces the empirical findings. Following this, Section 5 identifies new predictions arising from the model and verifies them. Then, Sections 6 and 7 relate ST² to competitor models of the blink. Finally, Section 8 discusses how the approach relates to theories of temporal attention in general and presents concluding remarks.

THE ST² THEORY

2. Informal Description of Model

The following principles underlie ST². We consider these in turn in Sections 2.1–2.5.

1. *Two stages.* We postulate a cascaded parallel first stage of visual processing, followed by a serial second stage tied to WM encoding.
2. *Salience filter.* We assume a mechanism for attributing salience, which filters the set of competing items.
3. *Types–tokens.* We distinguish between the featural representation of an item (its type) and the episodic context in which it occurs (a token). In ST², associating a token with a type is the process of WM encoding.
4. *Transient attentional enhancement (TAE).* We postulate a mechanism that, in response to the detection of a salient stimulus, provides a temporally brief (but spatially specific) enhancement of type representations.
5. *Binding pool.* We introduce a neurally plausible means to associate types with tokens.

2.1. Two-Stage Theory

Two-stage theories arise throughout the recent study of attention (Broadbent, 1958; Chun & Potter, 1995; Shih & Sperling, 2002; Treisman & Gelade, 1980). These theories are distinguished by the processes they ascribe to each stage and the properties of the interface between the two stages (Driver, 2001). The particular two-stage theory that ST² uses is best explained by considering the stages in turn.

¹ Two such examples are repetition blindness, which obtains when two instances of the same item arise in close temporal proximity (Kanwisher, 1987), and Lag 1 sparing, which shows that when items are presented in close temporal proximity, order is not easily retrievable (see Section 3.2 and the *Lag 1 Sparing* subsection of Section 3.3).

Stage 1

In the ST² model, Stage 1 implements standard visual processing: from visual feature extraction to semantic categorization. It is important to note that (modulo constraints arising from stimulus presentation²) Stage 1 is parallel. Thus, multiple items can be concurrently processed, with little interference between them. This is illustrated in Figure 1, in which progression through the stage is depicted as a vertical pass through a series of processing levels (horizontal bands).

In addition, processing is cascaded across the levels of Stage 1. This is illustrated in Figure 1 by depicting item representations as rectangles. Thus, item representations are “smeared” across a number of levels at any instant, with decreasing strength to the extremes of the rectangle. This is consistent with neural activation dynamics, which build up gradually and persist (in a decaying form) beyond the activation peak. A consequence of this cascaded processing is that multiple items can be coactive at the same level, even if they are presented to the system one after the other.

Effectively, each level of Stage 1 acts as a very short-term memory, the duration of which is governed by the speed of activation dynamics. Indeed, early levels of Stage 1 could be related to iconic memory (Coltheart, 1983; Shih & Sperling, 2002). Thus, representations persist for hundreds of milliseconds unless overwritten, as can arise from masking (Keysers & Perrett, 2002). However, these memories are indeed very short term. Thus, in the absence of top-down enhancement, Stage 1 representations are subject to rapid decay as well as being vulnerable to overwriting.

Stage 2

For an item to obtain a more durable representation, it has to make it through to Stage 2. That is, Stage 2 is the entrance to WM,

with a probable role for conscious perception in this process. In addition, in contrast to Stage 1, the second stage imposes sequentiality constraints. These arise because the system attempts to associate items with discrete episodic contexts. We consider the mechanism by which such associations are built when we discuss the types–tokens distinction.

2.2. Saliency Filter

We use the term *saliency filter* to include all mechanisms that provide top-down selective enhancement, whether driven by long-term cognitive goals or emotional significance. Such mechanisms can operate at many points in the Stage 1 processing pathway. In particular, task set modulates the response of neurons throughout the visual processing hierarchy (Desimone & Duncan, 1995).

We depict this mechanism (see Figure 1) as a processing layer that selectively emphasizes certain channels and de-emphasizes others. This approach is consistent with the task demand systems used in connectionist models (Cohen & Huston, 1994; McClelland, 1993). However, Figure 1 is not intended to indicate that the focus of the saliency filter is permanently fixed. Rather, the filter is being constantly adjusted, both horizontally (in terms of Stage 1 channels emphasized and de-emphasized) and vertically (in terms of level of operation). Indeed, it is likely to operate in graded fashion at many levels at the same time.

It is through the saliency filter that cognitive control (Botvinick, Braver, Barch, Carter, & Cohen, 2001) enforces the task set (Monsell, 2003). Thus, the saliency filter enhances task-relevant items, enabling them to progress into Stage 2. In contrast, the filter typically ensures that task-irrelevant items do not reach Stage 2.

2.3. Types–Tokens

The types–tokens distinction has been useful in explaining a number of temporal attention phenomena (Chun, 1997; Kanwisher, 1987, 1991), and it plays a central role in the ST² model. Before we come to this though, we define terms. *Types* encompass all featural properties of an item. For example, the type representation of the letter *K* would contain (a) semantic features, for example, that it is in the category of letters and it follows *J* in the alphabet, and (b) visual features, for example, its shape, the angled line segments that comprise it, and the color in which it appears.

In contrast, a *token* is a compact encoding of instance-specific (or episodic) information. Thus, a token indicates that a particular type has occurred and also where relative in time to other items it occurred. Note that because of our focus, *where* is interpreted in temporal (rather than spatial) terms. Our hypothesis is that although type representations are activated when detecting and identifying an item, it is unlikely that it is the type that is held active when the item is maintained in WM (see Section 6.1 for a discussion of how this impinges on active memory models). Rather, we argue that there is a transition from full type activation to compact token encoding as an item passes to a more durable representation. Thus, in the ST² model, WM encoding is exactly the process of associating (or binding) a token to a currently active type; we also

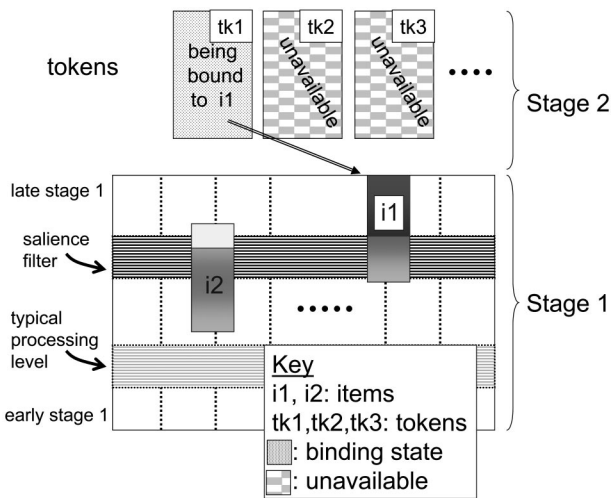


Figure 1. Depiction of the simultaneous type, serial token model’s processing. Items 1 and 2 are progressing simultaneously and independently through Stage 1. The saliency filter is enhancing the i1 channel and suppressing the i2 channel, shown as a change in the darkness of the rectangles depicting items. The system is in the process of binding a Stage 1 item or type into working memory. A color version of this figure is available in the online supplemental data for this article.

² For example, spatial overlap of input stimuli, as arises in AB experiments, can generate interference at the earliest levels of Stage 1; see the *Early Visual Processing: Masking Layer* subsection of Section 4.3.

use the term *tokenization* to describe this process. In this sense, once bound, tokens act as “pointers” from which the corresponding type can be regenerated when required.

Furthermore, once a token is bound to a type, in the absence of continued bottom-up support, activation of that type can “safely” be dissipated. In terms of Figure 1, tokenization frees up the Stage 1 channel of the type that has been tokenized. The fact that a second instance of a previously encoded item can be perceived at all suggests such a type dissipation (although, repetition blindness [Kanwisher, 1987] suggests that this liberation of resources can incur costs). That is, if WM maintenance were to commit the neural resources of a type, then those resources would not be available to detect a repetition of that type. Indeed, such an approach would also have difficulty retaining distinct items with common features. The alternative to this is that the brain holds duplicates of each type, with the number of duplicates required being bounded, at the least, by the capacity of WM.³ However, this would (a) require a complex type handover mechanism and (b) be exceptionally expensive in respect of neural resources.

Figure 1 depicts a snapshot of the tokenization process. Tokens are made available in sequence, and the model is shown in a state in which the first token is available, and all other tokens are unavailable. Item i1 has reached the end of Stage 1 and, thus, is fully (type) processed. In particular, i1 has largely passed the salience filter and, thus, has been identified as suitable for WM encoding.

Tokenization is initiated when an item passes the salience filter. Thus, Figure 1 shows a state in which i1 is being bound to the available token. Once this tokenization is complete, the system enters a state such as that shown in Figure 2. Once tk1 has been bound, a handover takes place and tk2 becomes available; that is, tokens are serial.

However, a neurobiological realization of tokenization would necessarily be complex; as a result, tokenization is not an instantaneous process. This has important implications because the binding of tokens to types is promiscuous in nature. Although Stage 1 can represent multiple items concurrently, the tokenization process

binds the currently available token to *all* types active after the salience filter. For this reason, it is important that during token binding, subsequent items be prevented from becoming strongly active. Thus, to protect the integrity of ongoing tokenization, attentional resources are withheld to prevent interference from other Stage 1 items. In ST² terms, the attentional resource is the TAE; see Section 2.4.

In RSVP, the task is to detect and/or identify the salient items in the sequence. In order to do this, participants must segment the presentation stream into discrete components, with key components being the salient items. The token system allocates discrete episodic contexts consistent with such segmentation. If a sequence is presented at normal speeds—that is, a good deal slower than in RSVP—each item encoded into WM is bound to a separate token. As a result, serial order can be retrieved; that is, the item bound to Token 1 is first, the item bound to Token 2 is second, and so on.

In addition, the ST² approach can encode a rich variety of type–token associations. First, the system enables multiple instances of a type to be recorded. That is, it can reach a state in which multiple distinct tokens are associated with the same type. It is also possible for two tokens associated with the same type to be adjacent in time. This will arise if an item is repeated with a sufficient gap in time between the two instances to ensure that the first tokenization has completed before the second starts.

Second, in extreme situations, the system can reach states in which two types are bound to the same token. This represents a state in which the system has failed to encode a unique episodic context to the occurrence of each of the types and plays an important role in our modeling of the AB; see the *Lag 1 Sparing* subsection of Section 3.3. Finally, inherent in the types–tokens approach is a justification for why the two-stage theory has the characteristics it does:

1. Stage 1 is parallel, because multiple distinct types can be simultaneously represented;
2. Stage 1 is subject to rapid forgetting, because active types decay;
3. Stage 2 ensures a more durable representation, because token binding compactly encodes a recognition event, which can be maintained without committing type representation resources; and
4. Stage 2 is serial, because it seeks to provide an unambiguous binding of types to discrete episodic contexts in the setting of a promiscuous tokenization process.

2.4. TAE

An item passing the salience filter in a strongly active form, as Item 1 has done in Figure 1, initiates an attentional enhancement, which elevates activation across the later levels of Stage 1. In this sense, the mechanism is exogenous in character, being instigated

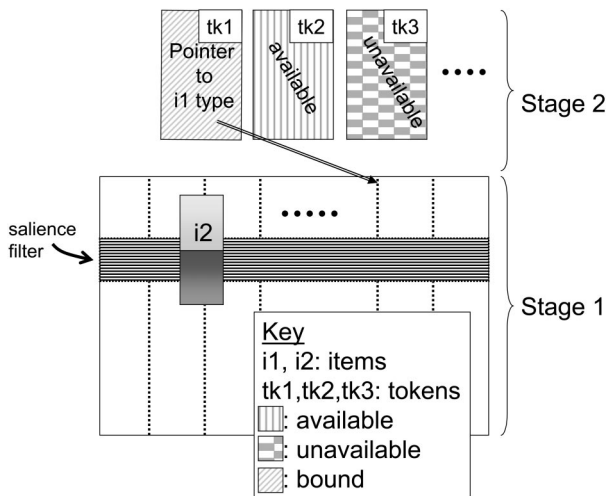


Figure 2. Item 1 bound to Token 1. A color version of this figure is available in the online supplemental data for this article.

³ The capacity of WM is a lower bound on the number of duplicates because it classically measures the number of distinct items that can be stored; however, the number of repetitions of the same item is likely to be a good deal bigger.

by the occurrence of an environmental stimulus. This mechanism highlights a very brief window of time and space that is particularly salient. The mechanism has a number of characteristics.

1. The TAE enhances the activation level of salient items, aiding their encoding into WM. This is particularly important when considering demanding stimulus environments (such as RSVP) because, unaided, fleeting representations have insufficient bottom-up activation to facilitate encoding. That is, attentional enhancement plays a key role in facilitating type–token binding.
2. The TAE has two important characteristics: (a) The attentional enhancement is brief (a pulse of around 100 ms), and (b) it only fires once per tokenization. Thus, detection of a salient item fires the TAE, but then it is maintained offline until this tokenization is complete. Unavailability of the TAE protects ongoing tokenization from the intrusion of newly arriving items, in an attempt to ensure discrete allocation of episodic contexts (i.e., tokens).
3. We hypothesize that the TAE is spatially specific but is not featurally selective in its enhancement. That is, it is initiated by detection of features characterizing the task set, but the effect of the TAE is not restricted to those features.

A TAE is suggested by RSVP data, particularly, one of the trademarks of the AB, Lag 1 sparing (see the *Lag 1 Sparing* subsection of Section 3.3). However, the TAE is also supported by other psychophysics research. For example, Nakayama and Mackeben (1989) described two forms of covert visuospatial attention. One was a sustained component that was slow to deploy. The other was a transient component with behavioral effects beginning 50 ms after a task relevant cue, but then fading within 150 ms. In their paradigm, these two forms of attention sum.

In addition, Muller and Rabbitt (1989) found elevated performance when cues preceded a fleeting target by 100 or 175 ms. They made the distinction between peripheral cues that appeared in the same location as the target they cued and central cues (e.g., arrows), which appeared at fixation and directed attention to the target. They found the same pattern as Nakayama and Mackeben (1989); that is, peripheral cues first evoked a transient pattern of improved accuracy over stimulus onset asynchronies (SOAs) of 100 and 175 ms, which then fell to a baseline defined by the sustained component of attention. Recent work exploring transient attention has identified similar effects (Kristjansson, Mackeben, & Nakayama, 2001; Kristjansson & Nakayama, 2003).

Deploying attention to any stimulus that is fleeting requires a mechanism with a rapid onset, as we describe here. This situation would, for example, arise when attempting to identify objects that are intermittently occluded. Furthermore, its rapid onset is well suited to deploying a brief episode of attention during a single saccade.

2.5. The Binding Pool

Up to this point, we have drawn arrows linking type channels and tokens as required; however this simplicity of depiction belies

the technical difficulties involved. Indeed, from a neural perspective, token binding is the most challenging aspect of our theory. This section introduces a suitable mechanism: the binding pool. The following are requirements for this mechanism.

1. It needs to be able to represent arbitrary type–token bindings. In particular, it should be possible to bind multiple tokens to the same type; see Section 2.3.
2. Token bindings need to be set up in real time and on demand, that is, when a salient item is detected.
3. We require that the mechanism encode bindings efficiently and in a neurally feasible manner (see Section 4.6). It is easy to describe strategies that are inefficient. For example, the number of neurons needed to fully characterize a type is likely to be very large. Therefore, an approach requiring the full neural profile of each type to be duplicated (in order, say, to encode repetitions) would be extremely wasteful and is unlikely to have arisen from evolution.
4. Following the previous point, it is also necessary to identify how the approach would scale to a brain-level realization

Hebbian learning is selective for synapses that are conjunctively (pre- and postsynaptically) active (Hebb, 1949; O'Reilly & Munakata, 2000) and, thus, seems to be ideal for binding types to tokens. However, the time scale required for these changes, which occur in vitro over the course of minutes, suggests that it is unsuitable. This is especially true in the temporal attention context in which multiple WM encodings can occur in a few hundred milliseconds.

Another candidate mechanism is short-term synaptic modification, whereby the responsiveness of a synapse is a complex function of its recent activity (Liaw & Berger, 1996; Pantic, Torres, Kappen, & Gielen, 2002). This mechanism is certainly rapid enough; however it occurs without postsynaptic specificity. Thus, coactivation of a type and a token would cause nonselective potentiation from a token to all types or vice versa. Consequently, information about which type co-occurred with which token would not exist.

We are left with an account based on sustained activity, which is probably the standard explanation for WM (Cowan, 2001). In particular, a number of buffer models exist that posit sustained activity at various levels of representation (Amit, Bernacchia, & Yakovlev, 2003). Typically though, these models do not fulfill the requirements identified in Section 1.2 (see Section 6.1 for a detailed justification). For example, objects would be stored by sustained activity in the same (type) nodes used to recognize those items, making it impossible to detect repetitions, among other limitations. Note that this same problem would arise with an approach based on temporal synchrony (Hummel & Biederman, 1992; von der Malsburg, 1981). Further, pure active memory models typically leave unexplained the means by which stored memories are accessed. This is in contrast to types–tokens accounts, in which active tokens are a localized focus through which encoded types can be regenerated.

In response, we propose a significant modification of the active memory approach. In ST^2 , a dedicated pool of binding nodes (see Figure 3) forms a self-sustaining pattern of activity during encoding to store the coincidence of types and tokens. Binding pool units are selective for specific types and tokens, with each type and token being coded for by overlapping subsets of the pool. Figure 4 shows a binding pool state arising when binding Type i1 to Token tk1. As is typical of the approach, some binding pool units code neither i1 nor tk1, and some code just tk1, some just i1, and others both (i.e., the conjunction of) tk1 and i1. It is these conjunctive units that bind i1 to tk1 and that become (durable) active representations. Thus, the sort of binding link we have been depicting as an arrow between a token and a type is effectively piggybacked on top of the active binding pool representation.

The binding pool satisfies the requirement that WM does not commit type representation space, thus allowing a type to be processed even if it is already present in the memory set. In Section 4.6, we discuss an implementation that clarifies how binding pool units are realized as neural circuits and the connectivity into and out of the pool. In this implementation, the binding pool contains localist representations of every possible combination of target type and token, similar to Deco, Rolls, and Horowitz’s (2004) approach to binding items to spatial locations.

In scaling such localist approaches to brain-level implementations, there is an inherent problem of combinatorial explosion in the size of the pool of neurons required to represent every possible combination (in our case, number of tokens \times number of types nodes). The use of a distributed representation within the binding pool solves this problem. A small pool of nodes can store a compact representation of all possible combinations of types and token (Wyble & Bowman, 2006). However, with a distributed approach, there is overlap between the representations of different token–type bindings, with the weights into and out of the binding pool governing the level of overlap (indeed, it is this overlap that leads to a more efficient encoding). As a result, the distributed approach produces a falloff in retrieval fidelity as memory load

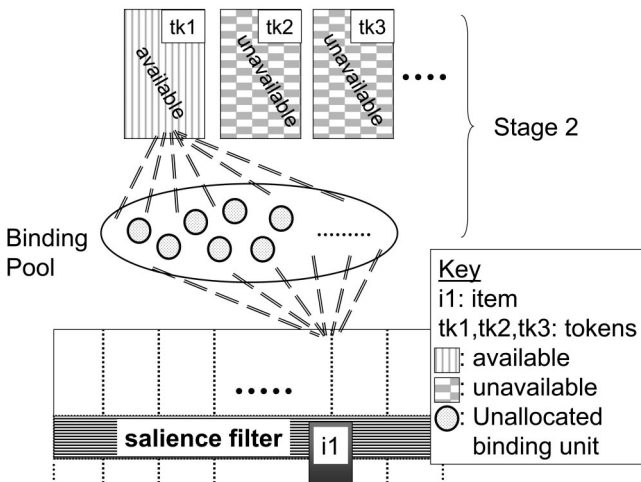


Figure 3. The simultaneous type, serial token model’s binding pool, before any tokenization. Note that the connectivity shown would be duplicated across type and token layers. A color version of this figure is available in the online supplemental data for this article.

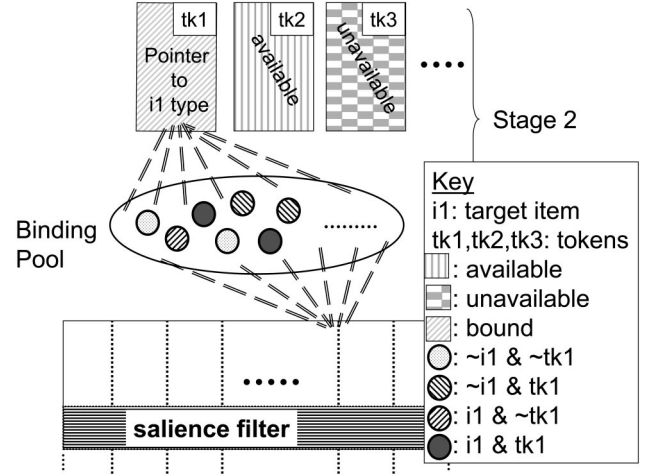


Figure 4. The simultaneous type, serial token model’s binding pool in a state arising from coactivation of Token 1 and the Item 1 type. Fill denotes what the binding unit codes: neither i1 nor tk1, just tk1, just i1, or both i1 and tk1. A color version of this figure is available in the online supplemental data for this article.

increases, replicating Wilken and Ma’s (2004) experimental results. Thus, although for simplicity of modeling and presentation we use a localist binding pool in this article, we believe that distributed binding pools would scale.

2.6. Discussion

The ST^2 model satisfies the five functional requirements of temporal attention highlighted in Section 1.2.

1. *Salience detection.* ST^2 ’s salience filter detects salient stimuli when briefly presented in visual environments containing (temporally) competing stimuli.
2. *Sustain representations.* The TAE has the capacity to sustain fleeting, but salient, representations of visual stimuli, enabling WM encoding.
3. *Ascribe episodic context.* ST^2 ’s types–tokens system enables discrete episodic contexts to be ascribed to items. Indeed, the allocation of such a token is the process of encoding into WM. In addition, tokens are allocated in sequence, enabling the order of perceived occurrence of items to be retrieved.
4. *WM maintenance.* Because multiple tokens are available, many items can be maintained together in WM.
5. *Repetitions.* Multiple tokens can be bound to the same type, enabling repetitions to be maintained in WM. Furthermore, because type resources are freed up once an item has been tokenized, those resources are available to perceive and encode further instances of the same item.

Finally, regarding the standard debate between early and late selection in attentional processing (Driver, 2001; Logan, 2004), we

are largely noncommittal, because configurations of the salience filter at early levels of Stage 1 would generate early selection, whereas configuration at late levels would generate late selection. This said, the salience filter configurations considered in this article would enforce late selection, as suggested by the characteristics of the AB (see Section 3.2). It is important though to realize that in contrast with many other prominent two-stage theories (e.g., Broadbent, 1958), it is not the filter that is the interface between Stage 1 and Stage 2 or (in an absolute sense) imposes seriality. In ST^2 , the salience filter certainly is a bottleneck, because it restricts what passes through it, but it does not, per se, enforce seriality.

THE EMPIRICAL LANDSCAPE

3. Temporal Attention Studies

We assess the ST^2 model against a set of archetypal temporal attention data focused on the AB.

3.1. The AB

Perhaps the most prominent RSVP paradigm is the AB (Raymond et al., 1992). In this task, two targets are placed in a sequence of items presented at a rate of around 10 items per second. We focus on the *letters-in-digits* task (Chun & Potter, 1995), in which participants must report the identity of two letter targets (T1 and T2) presented in a stream of digit distractors. Our reason for selecting this task is that certain aspects of other tasks would not be straightforward to model. In particular, the letters-in-digits task does not require binding across features, for example, between color and identity, as found in color marked tasks (such as in Maki, Couture, Frigen, & Lien, 1997). Furthermore, the task can be argued to yield a pure test of the blink, because there is no task switch between the T1 and T2 tasks (Chun & Potter, 2000).

3.2. Findings To Be Reproduced

The specific findings that we reproduce are as follows:

1. *The basic blink.* A typical AB serial-position curve, which arises from the letters-in-digits task, is shown in Figure 5. Points to note are
 - a. the blink is a 200-ms to 500-ms (approximately) interval post T1 onset in which performance on T2, conditional on correct report of T1 (i.e. $T2|T1$), is significantly below baseline;
 - b. generally, the blink has a sharper onset than offset; and
 - c. if T2 immediately follows T1, it is reported at baseline levels, which is described as *Lag 1 sparing*.

Of particular note is Lag 1 sparing, which seems counterintuitive, because one may believe that limited attentional resources are maximally engaged in processing T1 when T2 “arrives” at Lag 1.

2. *T1 + 1 and T1 + 2 blank.* The blink was attenuated when a blank was placed in the T1 + 1 position, but not when it was placed at T1 + 2 (Chun & Potter, 1995; Raymond et al., 1992); see Figure 5.

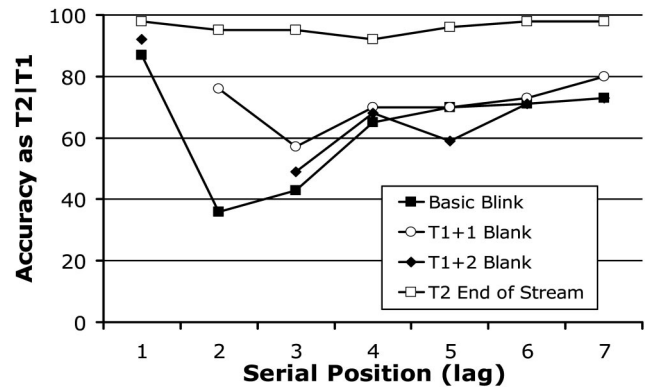


Figure 5. Serial position curves for letters-in-digits tasks, with a 100-ms stimulus onset asynchrony. Conditions were no blank (i.e., the basic blink), a blank at Target 1 (T1) + 1, a blank at T1 + 2, and Target 2 (T2) as end of stream. Because of the insertion of blanks, some data points do not exist. The basic blink, T1 + 1 blank, and T1 + 2 blank data are from “A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation,” by M. M. Chun and M. C. Potter, 1995, *Journal of Experimental Psychology: Human Perception and Performance*, 21, p. 114, Figure 4. Copyright 1995 by the American Psychological Association. Adapted with permission of the author. The T2 end of stream data are from “Beyond the Attentional Blink: Visual Masking by Object Substitution,” by B. Giesbrecht and V. Di Lollo, 1998, *Journal of Experimental Psychology: Human Perception and Performance*, 24, p. 1457, Figure 2B. Copyright 1998 by the American Psychological Association. Adapted with permission of the author.

3. *T2 unmasked.* The blink was significantly attenuated if T2 was the last item in the stream (Giesbrecht & Di Lollo, 1998); see Figure 5. In addition, we suggest that placing blanks after T2 will attenuate the blink.
4. *T1 performance.* T1 performance was reduced at Lag 1; see Figure 6.
5. *Swaps.* It is possible that both T1 and T2 are identified, but they are perceived in the wrong order. In Chun and Potter (1995), such T1–T2 swaps were maximal at Lag 1 and effectively disappeared by Lag 3; see Figure 6. In fact, at Lag 1, temporal order judgment was only a little above chance.
6. *Stage of the AB bottleneck.* A number of studies indicate that the blink results from a late stage bottleneck.
 - a. *Priming.* Shapiro, Driver, Ward, and Sorensen (1997) considered AB paradigms with three targets. In one task, T2 was an uppercase letter and T3 was lowercase. The basic finding was that when a T2 was missed, correct report of T3 was higher if the two targets had the same identity. In the second experiment, the stream items were words, and it was discovered that missed T2s semantically primed T3.
 - b. *T2 breakthrough effects.* High salience T2s breakthrough the blink. For example, substantial blink attenuation was shown when personal names were used as the T2 (Shapiro, Caldwell, & Sorensen, 1997). In addition, emotionally salient T2s attenuated the blink (Anderson & Phelps, 2001).

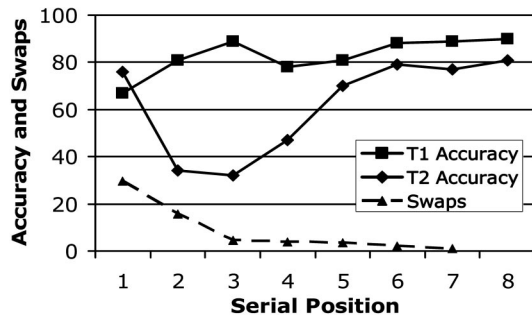


Figure 6. Serial position curves of pure Target 1 (T1) performance, pure Target 2 (T2) performance, and percentage of T1–T2 order inversions (swaps) for letters-in-digits task. From “A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation,” by M. M. Chun and M. C. Potter, 1995, *Journal of Experimental Psychology: Human Perception and Performance*, 21, p. 112, Table 1, and p. 119, Figure 8. Copyright 1995 by the American Psychological Association. Adapted with permission of the author.

c. *Event-related potentials (ERPs)*. Early visual components of the ERP waveform (the N1 and P1) elicited by T2s have been shown to be comparable whether T2 was missed or seen. In addition, semantic effects (N400) of a T2 word were also present when that item was missed. In contrast, a component associated with WM update (the P3) was reduced when a T2 was presented during the blink compared with when it was presented outside that period (Luck, Vogel, & Shapiro, 1996; Vogel, Luck, & Shapiro, 1998).

These investigations suggest that even when a T2 is missed during the blink, it is nonetheless extensively processed, in respect of both visual and semantic features, allowing priming and breakthrough. Furthermore, the ERP P3 data suggest that the blink bottleneck is located at the point of encoding into WM.

7. *Priming from T1 + 1 distractors*. Chua, Goh, and Hon (2001) found that performance on a color marked T2 letter was improved if the same letter was presented as a distractor in the T1 + 1 position, suggesting that distractors following the T1 prime later T2s.

3.3. The ST^2 Model and the AB

Before we move to introducing our neural implementation, we consider how the AB fits with our informal formulation of the ST^2 model.

Fleeting Representations and Bottom-Up Trace Strength

In the context of RSVP, we argue that the fleetingness of visual stimulus representations arises from masking. That is, in a basic blink stream, each item (apart from the first) acts as a mask for the item that precedes it.⁴ As a result, masking attenuates sensory traces of items. Accordingly, we distinguish between *weak* and *strong sensory traces*, which arise when an item is masked and

unmasked, respectively. The key situation in which items are unmasked is when they are followed by a blank.

In a manner that we clarify shortly, strong traces can be more easily encoded into WM than weak traces. It is for this reason that the blink is attenuated in the T1 + 1 blank and T2 as end of stream conditions; see Figure 5. This interpretation is consistent with a number of studies that have explored the effects of low-level masking in the AB (e.g., Seiffert & Di Lollo, 1997).

Why the ST^2 Model Blinks

In terms of the ST^2 model, the blink occurs because although Stage 2 is occupied encoding T1, T2 is decaying at Stage 1 and may have decayed completely by the time Stage 2 is again free. Specifically, the blink arises because the tokenization process is overloaded, in the sense that targets arrive so rapidly that T1 is still being tokenized when T2 arrives. As a result, the system sacrifices tokenization of T2 (by taking the TAE offline) in order to prevent it from interfering with the tokenization of T1. That is, the blink is a mechanism to avoid spurious and confused binding of T2 to a token that is in the process of being allocated to T1.

Lag 1 Sparing

It may seem that Lag 1 sparing is a problem for this explanation of the blink, because at Lag 1, T2/T1 performance is high even though the T1 and T2 tokenization processes maximally overlap in time. However, Lag 1 sparing does not come free of cost. First, as discussed above, T1 accuracy is degraded at Lag 1, suggesting that the competing T1 and T2 tokenizations can resolve in favor of T2 and at the expense of T1. Second, as again previously highlighted, swaps are very high at Lag 1.

In ST^2 terms, in the Lag 1 case, when the TAE fires in response to T1 detection, T2 is sufficiently advanced in Stage 2 that it benefits from that attentional enhancement (indeed, as suggested by Potter, Staub, & O’Conner, 2002, it is likely that the T2 benefits more than the T1). As a result, both T1 and T2 become strongly active together in the postsaliency filter levels of Stage 1, which yields the degenerate situation in which two types are bound to the same token. Loss of temporal order information is characteristic of such a failure of the tokenization process.

More generally then, we argue that because of the rapid arrival of fleeting representations, the system is unable to extract all the information available in the stimulus stream. Thus, it is forced to trade extraction of one information-bearing dimension off against extraction of another. Indeed, the transition from Lag 1 to Lag 2 represents an adjustment of this trade-off, from extraction of information about both targets at the expense of confused binding (at Lag 1) to extraction of unequivocal T1 information at the expense of T2 information (at Lags 2, 3, and 4).

⁴ Although, masking in AB experiments is weaker than that arising in classic psychophysics masking experiments (Keysers & Perrett, 2002). In particular, in AB tasks items are presented for around 100 ms, which is a good deal longer than in classic masking studies.

NEURAL REALIZATION

4. The Neural ST² Theory

This section describes a neural realization of the ST² model (called neural-ST²). We begin, in Section 4.1, by introducing a set of neural principles. Then, in Sections 4.2–4.7, we describe the neural realization. In Section 4.8, we show the data the model reproduces, and finally, in Section 4.9, we relate the model to neurophysiology.

4.1. Neural Representation Scheme

We first describe activation equations and then highlight key neural mechanisms. However, our description of the model can be comprehended without understanding these mathematical details.

Simple Elements

Activation functions of neural units are simple combinations of bias, excitation, inhibition, and leak currents. Connections between elements are excitatory or inhibitory and are not modifiable. The membrane potential update function is

$$\begin{aligned} MP_{(i,j,t)} = & MP_{(i,j,t-1)} + DT_VM_{(j)} \times \\ & [(Bias_{(j)} + Excite_{(i,j,t-1)}) \times (EE_{(j)} - MP_{(i,j,t-1)}) + \\ & Inhib_{(i,j,t-1)} \times (EI_{(j)} - MP_{(i,j,t-1)}) + \\ & Leak_{(j)} \times (EL_{(j)} - MP_{(i,j,t-1)})], \quad (1) \end{aligned}$$

where $MP_{(i,j,t)}$ is the membrane potential for node i of layer j at time t .

The membrane potential is a measure of the extent to which a node is excited. Thus, when MP is high, the node is very excited, and when it is low, it is unexcited. DT_VM is a constant determining the rate of change of a node. $Bias$ is a constant input current, which regulates the basic excitability of a node; thus, the higher the bias, the more easy it is to push the membrane potential over its output threshold (see Equation 2). $Excite$ and $Inhib$ are weighted summations of all excitatory and inhibitory inputs respectively. $Leak$ is a leak current, which determines how quickly the node decays back to its resting level.

EE , EI , and EL are reversal potentials for excitatory, inhibitory, and leak currents, respectively. Excitation and bias push the node up to the value of EE . In contrast, inhibition and leak push the node down to EI and EL respectively. Thus, EE bounds the membrane potential from above and EI (which is set to be smaller than EL) bounds the potential from below. In addition, EL sets the resting level, because the node will stabilize at this value if it has no excitatory or inhibitory input for a period of time. The output mapping is a sigmoid function, which bounds a node's output to the range $[0, 1]$:

$$Out_{(i,j,t)} = \frac{[MP_{(i,j,t)} - \theta_{(j)}]_+ \times \gamma_{(j)}}{([MP_{(i,j,t)} - \theta_{(j)}]_+ \times \gamma_{(j)}) + 1}, \quad (2)$$

where

$$[x]_+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Thus, if the membrane potential is below the threshold θ then the node outputs zero. However, the output rises rapidly as the membrane potential becomes bigger than θ and then saturates at large values. How rapidly output approaches the saturation limit of 1 is governed by the scaling parameter, γ . These equations are similar to those found in O'Reilly and Munakata's (2000) biologically plausible neural networks and can be related to Hodgkin–Huxley equations (Koch, 1999).

Key Neural Mechanisms

We now consider three mechanisms of which we make liberal use.

Self-excitation. In the absence of excitatory input, leakage ensures that a node's activation decays exponentially back to resting levels. However, we need to model a more temporally sustained maintenance of activation, as found for example, in activation based memory (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Davelaar & Usher, 2004). In localist models, this effect is obtained by including an excitatory link from the node back to itself (see Figure 7a). Such a self-excitatory node can be viewed as an abstract representation of an assembly of biological neurons, which sustains activation across the assembly through recurrent interactions between constituent neurons.

Off nodes. It can also become necessary to break a self-sustaining cycle. A simple way to do this is to use an inhibitory interneuron (i.e., an off node; see Figure 7b), which builds up in response to activation of its associated self-sustaining node, and once the inhibitory interneuron crosses its output threshold, it suppresses the self-sustaining node. In fact, neurobiologically speaking, off nodes could be realized in a number of ways. Off nodes could be instantiated as an intracellular process; for example, a calcium activated after-hyperpolarization potassium conductance current is found in most neurons. This mechanism reduces excitability following periods of strong activation (Bal & McCormick, 1993; Hotson & Prince, 1980). Alternatively, off nodes

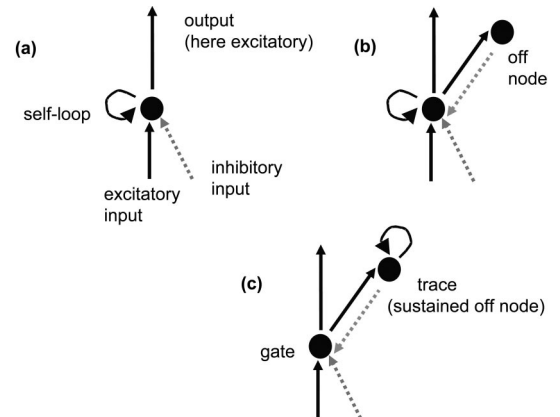


Figure 7. a: A basic node with a self-excitatory link. b: A basic node with self-excitation and an inhibitory interneuron. c: A basic (non-self-sustaining) node with a self-excitatory off node. Solid links are excitatory, and dashed links are inhibitory.

could be viewed as a localist implementation of a layer of inhibitory interneurons (O'Reilly & Munakata, 2000).

Gate-trace circuit. This neural structure revises the off node circuit shown in Figure 7b by moving the self-loop to the off node, as shown in Figure 7c. We call this a gate-trace pair. As a result, the (off) trace node can be activated only through the (on) gate node, and once activated, the trace node stays active. (The fact that the trace node stays active is a major difference to the previously discussed [off node] circuit, in which, once the off node fires, no activity is maintained in the circuit.) Typical behavior of such a circuit is depicted in Figure 8. Figure 8A shows the circuit in its initial state. Figure 8B shows the state reached when the circuit has been externally excited through the gate node. As the gate node becomes activated it excites its trace node, which, when strongly active, suppresses the gate. Eventually the circuit settles into a new stable state with a strongly active trace node and a suppressed gate node (as shown in Figure 8C). Thus, the trace node is an active memory, which is accessed through the corresponding gate node. In addition, when in its sustaining state (see Figure 8C), the gate node can be reactivated by suppressing the trace node, thus freeing the gate to be biased by extrinsic influences. This method can, for example, be used to retrieve from the active memory. A layer of such gate-trace pairs amounts to a *gated active memory*. A gate-trace pair provides sustained activation without tying up the whole layer. For example, competition (via lateral inhibition) among gate nodes can continue in the presence of sustained trace activation. A simple layer of competing self-sustaining nodes does not behave in this way: Once a node in the layer is in a self-sustaining state, there is no future opportunity for competition.

4.2. Overview of Neural-ST²

The neural-ST² model is depicted in Figure 9. ST²'s first stage is realized by the bottom four layers (input, masking, item, and task filtered), which model the extraction of types. The connectivity (which is one-to-one between layers and has no lateral inhibition at the task filtered layer [TFL] and item layer) ensures that Stage 1 can represent multiple items simultaneously. The

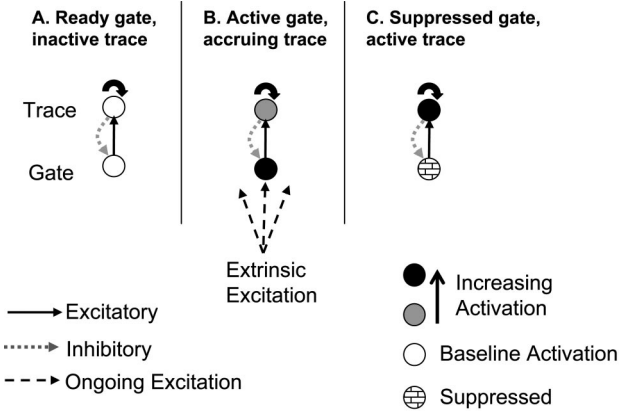


Figure 8. Activation dynamics of a gate-trace circuit, which is shown in three consecutive states. A: A stable (resting) ready state. B: A dynamic state in response to extrinsic excitation. C: A stable active maintenance state.

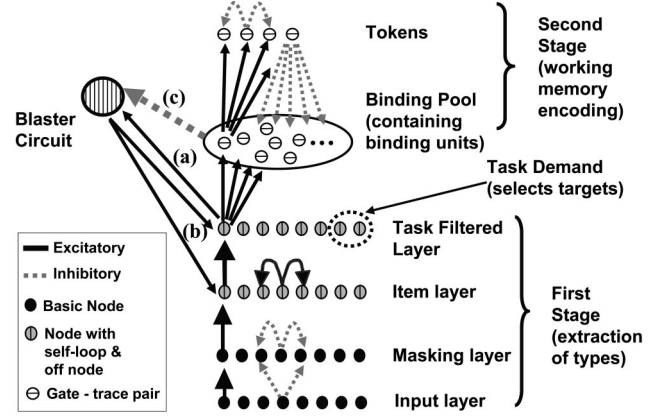


Figure 9. Structure of the neural simultaneous type, serial token model. Connection patterns between layers are depicted only for some nodes but apply for the entire layer.

second stage is realized by the WM layer (comprising tokens) and their interaction (through the binding pool) with the TFL.

Furthermore, the second stage is serial in the sense that only one token is available at any time. Finally, ST²'s TAE, which we realize with a mechanism called *the blaster*, is implemented as a neural circuit, with the projections into the circuit controlling the TAE's availability. We now move to a more detailed description of these mechanisms. First, in Section 4.3, we review the first stage. Then, we describe the second stage in Section 4.4. Following this, in Section 4.5, we consider neural-ST²'s TAE, and then, in Section 4.6, we discuss the binding pool. Finally, in Section 4.7, we present a brief summary of the neural realization.

4.3. Neural-ST²'s First Stage

Input

The RSVP stream is modeled by presenting a sequence of patterns to the input layer. Each pattern models an item in the sequence and is presented for 20 time steps. Trials are 550 time steps, each step simulating 5 ms of time. Streams contain two targets and 16 distractors, with T1 always the seventh item and the two targets separated by 0–7 distractors. T1 and T2 strengths are varied systematically over the range of .448 to .0604 (in steps of 0.012) for each lag. This variance simulates featural aspects of letters and digits, which make some of them easier to perceive. Distractor activation values were fixed at .50. For the sake of simplicity, the model does not explicitly represent the binding of visual features into complete items. Rather, at all layers, a single node represents an item presented to the model. The model is thus localist in nature. In addition, there is no explicit representation of spatial location in the model, although this could easily be added. We have abstracted from these details, because our main focus is higher up the visual processing pathway, that is, the point of WM encoding.

Early Visual Processing: Masking Layer

Input activation propagates over unidirectional connections through layers that abstractly represent steps of visual processing.

In RSVP streams, because stimuli are presented at the same spatial location, successive items are represented using the same neural tissue. As a result, items overwrite one another, generating masking. Such overwriting is abstractly modeled at the masking layer, which includes inhibitory interactions between items; see Figure 9. The masking effects generated by such inhibition are a key determinant of bottom-up trace strength; see the *Fleeting Representations and Bottom-Up Trace Strength* subsection of Section 3.3. Thus, activation traces are greatly weaker for masked as opposed to unmasked stimuli; see Figure 10. Our approach is consistent with single cell recordings (e.g., Keyser & Perrett, 2002; Kovacs, Vogels, & Orban, 1995; Rolls, Tovee, & Panzeri, 1999) that show strong amplitude differences in the firing rate of temporal cortex neurons to different types of masking in monkeys.

Masking is asymmetric; that is, backward masking is stronger than forward masking (Seiffert & Di Lollo, 1997). This asymmetry arises because feed-forward inhibition into the masking layer is stronger than lateral inhibition within the layer. Lateral inhibition creates symmetric suppression between masking layer nodes; that is, adjacent items in the stream suppress each other to a similar degree. In contrast, because feed-forward inhibition generates suppression only from an active input node (as set by the current input pattern), it only generates inhibition from stream items to immediately preceding items.

Type Processing: The Item Layer and the TFL

The final two layers in Stage 1 (see Figure 11) abstractly represent type-oriented processing, including extraction of conceptual features (Potter, 1993). The first of these, the item layer, supports a temporally sustained (equivalent to several hundred milliseconds), but decaying, representation of items (see Figure 10). Task set has no effect on this layer (i.e., distractors and targets

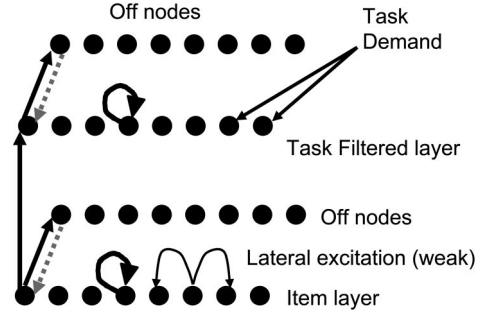


Figure 11. The type layers of the neural simultaneous type, serial token model. In all cases, apart from the task demand projections, the connectivity depicted for individual nodes extrapolates to all nodes in the layer. We show the constituent components of the task filtered and item layers, that is, self-loops and off node circuits. Thus, this figure expands out the gray filled nodes in Figure 9. Solid links are excitatory, and dashed links are inhibitory.

are treated identically), and similar items have weak excitatory connections (see Figure 11), along which an active item can prime similar items. Thus, even weak T2s, which are more likely to be missed, can prime future targets, as required by the late blink bottleneck hypothesis. Despite the fact that the item layer is localist, the brain may realize the same effect through an efficient distributed representation scheme, with priming arising as a by-product of the resulting overlapping item representations.

In contrast, at the TFL, task demand selectively excites target nodes while suppressing distractor nodes (see Figure 11). Thus, in respect of the ST^2 theory, the TFL enforces the salience filter, with, in this context, task instructions fully prescribing salience.

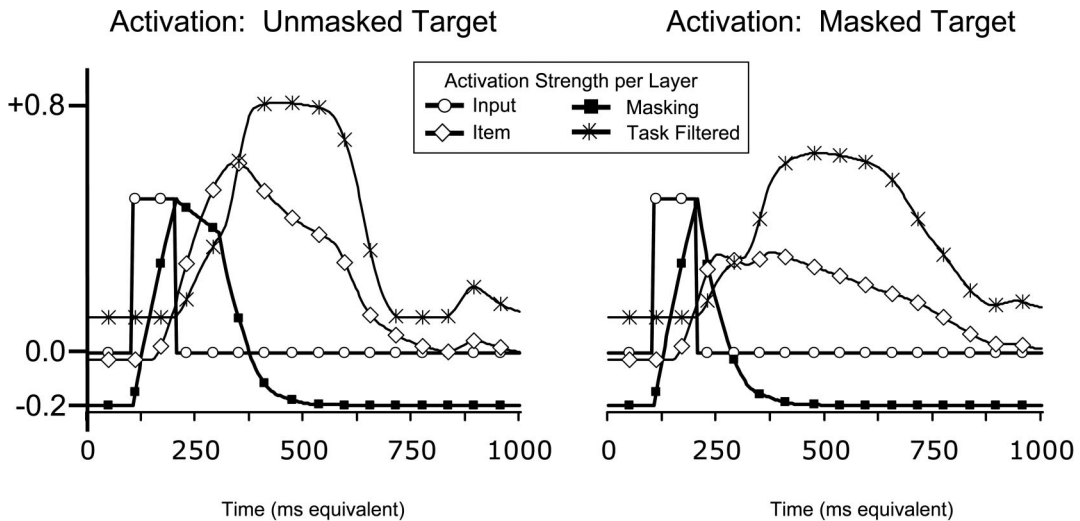


Figure 10. Input, masking, item, and task filtered layer activation traces for a target when it is masked and unmasked. Notice that masking layer traces have similar maximum amplitudes in both cases, but the unmasked trace is considerably longer than the masked trace. Broadly speaking, this trace length difference at the masking layer is turned into amplitude differences at the item and task filtered layers. In particular, the task filtered and item layer traces are shorter in the unmasked case, because tokenization completes more rapidly, and thus, off nodes cut in more quickly.

The TFL is the gateway to tokens. Excitatory connections from TFL items project to the binding pool, which, in turn, projects excitatory links to the gate nodes of tokens (see Figure 12). TFL and item nodes are weakly self-excitatory, yielding temporally sustained traces. However, these layers are also connected in a one-to-one fashion with off layers, which curtail activation traces; see *The Token System* subsection of Section 4.4.

4.4. *Neural-ST²'s Second Stage*

In accordance with the ST² theory, the neural-ST² first stage yields a decaying trace of the visual and semantic features of items (see Figure 10), which is parallel because different items can be simultaneously active at the item layer and TFL. In contrast, Stage 2 realizes encoding into WM, with sequentiality emerging from mechanisms that attempt to ensure that items are discretely bound into WM. We begin by discussing the token system. Then, we discuss the retrieval of items from tokens.

The Token System

Tokenization. During presentation of an RSVP stream, task-relevant Stage 1 items can activate a token, with this activation being relayed via the binding pool. This initiates a tokenization process, whereby binding pool units are allocated. The process of activating the token and allocating binding pool nodes is not instantaneous; it takes hundreds of milliseconds. However, the stronger the type trace, the faster it occurs.

The token mechanism is implemented as a layer of trace–gate pairs, each of which corresponds to a token; see Figure 12. That is, the rectangular tokens depicted in our informal exposition of the ST² theory (e.g., as shown at the top of Figure 2) are implemented as trace–gate circuits. Tokens have four states: *available*, *unavailable*, *binding*, and *bound*; see Figure 13. An available token has a minimally active gate node and an inactive trace node. While one token is available, others are unavailable because of lateral inhibition between the gates. A token in the process of binding has a highly active gate node, being driven (via the binding pool) from the first stage, and a trace node, which is steadily accruing activation. A bound token has a suppressed gate node receiving continuous inhibition from its active trace node.

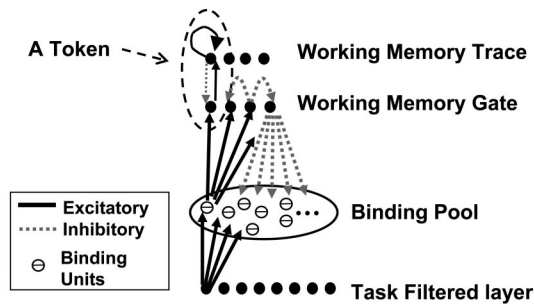


Figure 12. The token system. Connectivity shown for an individual node is repeated for each node in the given layer. As depicted, all binding pool units are unallocated.

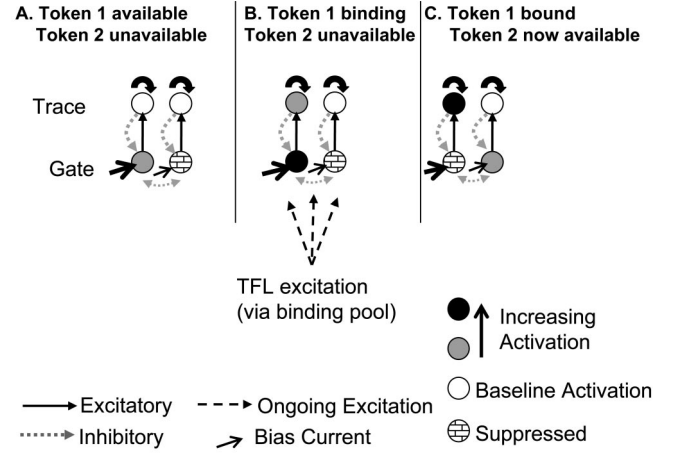


Figure 13. Token dynamics. Bias strength is indicated by point size of the arrowhead denoting the bias. TFL = task filtered layer.

While a token is being bound, binding pool units sitting between the token's gate node, and active TFL nodes are incrementally allocated. The rate at which these nodes are allocated is proportional to the strength of the individual traces in the TFL. The self-loop associated with each WM trace node ensures that once the corresponding WM gate is bound (as shown in Figure 13C for Token 1), the trace node will remain active indefinitely. Such sustained trace node activation is used later to determine which tokens were bound. Allocated binding pool nodes between the WM gate nodes and the TFL determine what item or items were encoded by those tokens.

Properties of the token system. WM gate nodes compete to become available at the beginning of a trial and also when a token has completed being bound. The winner is determined by an ordered pattern of biases applied to the WM gates, which determines which token is first, second, and so on; see Figure 13. In this way, neural-ST² ascribes discrete episodic contexts and encodes order when items are presented slowly enough.

It is possible that binding pool nodes allocated from different WM gates may project to the same TFL node. This reflects the fifth of the functional requirements we highlighted in Section 1.2, because it allows representation of multiple instances of the same item, that is, repetitions. It is also possible for binding pool units to be allocated between one token and many TFL nodes. That is, many items are encoded into WM, but order information is lost, which will yield order inversions at recall, that is, swaps.

TFL off nodes. Although it is a common phenomenon to fail to perceive a rapidly presented object, participants do not report extraneous instances of a single item presented within an RSVP stream. To ensure that neural-ST² also possesses this property, TFL nodes have their own off nodes; see Figure 11. These off nodes ensure that after a sustained period of activation, an item is inhibited to prevent the creation of spurious tokens. Without this mechanism, an item (at the TFL) could be active from one token to the next, being bound to multiple tokens.

Recall

At the conclusion of an RSVP stream, the model simulates recall. There are two phases to this process: (a) *identity retrieval* and (b) *order retrieval*.

Identity retrieval. The retrieval system first locates bound tokens (as indicated by active WM trace nodes) and then identifies binding units allocated to each token; if binding units are allocated, the type (or types) pointed to by those units is retrieved. Thus, for successful identity recall of a target, a portion of the binding pool must be allocated from any token to that target, and that token must have an active trace node. Multiple identities are retrieved if only one token has been bound, which, through allocated binding units, points to multiple types.

Order retrieval. If both targets were successfully encoded to either one or two tokens, the recall phase determines in which order they were perceived. This proceeds according to the following algorithm:

1. If either token (tk1 or tk2) is unambiguously bound (i.e., to a single target), then order is uniquely defined;⁵
2. otherwise there is a 50:50 chance of either report order.

Note that Alternative 1 handles the case in which both tokens are unambiguously bound, because the condition will trivially hold. In addition, Alternative 2 will be applied if all token bindings are ambiguous, whether there are one or two such bindings (both situations of which can occur at extreme presentation rates).

4.5. Neural-ST²'s TAE: The Blaster

As discussed in Section 3.2 and the *Lag 1 Sparing* subsection of Section 3.3, there are a number of pieces of AB data that support our TAE. First, Lag 1 sparing is suggestive of such a mechanism. Second, Chua et al. (2001) demonstrated that a distractor is a more effective prime of a T2 if it is preceded by a T1. Third, Potter et al. (2002) have shown that at very short SOAs (particularly less than 100 ms) T2 performance is in fact better than T1 performance. The implication is that perception of the T1 generates a brief window (around 150 ms) of enhanced processing that falls most strongly on the (possibly distractor) item in the T1 + 1 slot.

The blaster is depicted in Figure 9 (full implementation details can be found in Section A.2 of the Appendix). Above-threshold activity in any node of the TFL excites the blaster (through the projection labeled a in Figure 9). However, task demand foregrounding ensures that in practice only perception of a target can trigger the blaster, which in turn sends a powerful excitatory projection to all nodes in the TFL and item layer (through the projections labeled b in Figure 9). This causes a general excitation of all item and TFL nodes.

The blaster fires a burst of excitation and then rapidly shuts off. Furthermore, the shutoff is sufficiently strong to induce a refractory period, in which the blaster is unavailable. In addition, inhibition (through the projection labeled c in Figure 9) while binding units are being allocated maintains the blaster in this off state even when the refractory period is complete.⁶ Therefore, the general behavior of the blaster is a spike of excitation followed by a period

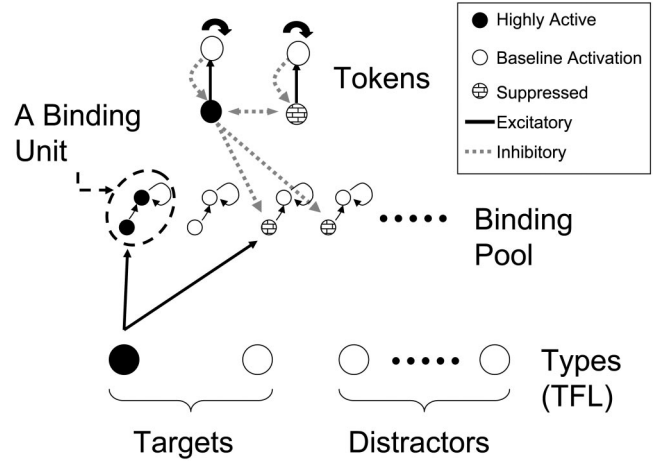


Figure 14. The binding pool in action. TFL = task filtered layer.

of inactivity until the completion of the current token binding. These dynamics protect the integrity of ongoing tokenization by limiting attentional resources that could cause binding intrusions.⁷

4.6. Neural-ST²'s Binding Pool

Each token provides a powerful inhibitory projection to three quarters of the binding pool, whereas type nodes in the TFL have excitatory projections to some of the units in the pool (see Figure 14). These projections ensure that for each combination of type and token, there will be one element of the binding pool that is (a) not inhibited by the token and (b) excited by the type. Thus, the pool can uniquely represent every combination of token and type as well as combinations of multiple types bound to a single token or multiple tokens bound from a single type.

The binding pool itself comprises gate–trace pairs that function much like the nodes making up each token. When confusion can arise, we use the terms *WM gate–trace* (or *token gate–trace*) to denote constituents of tokens and *binding gate–trace* to denote constituents of binding units. Within such a binding unit, the gate node is activated by input from the TFL and in turn excites the corresponding trace node, which slowly accrues activation. When the trace node reaches its threshold, it inhibits its corresponding gate node. The self-sustaining activity in the trace node serves to maintain the corresponding type–token binding.

⁵ That is, even if both targets are bound to the other token, then this ambiguity is resolved by the certainty of the binding to the unambiguously bound token.

⁶ Note that this inhibition is active only during a binding allocation and is not active while an existing allocation is being maintained. This effect is obtained because, as clarified in Section 4.6, each binding unit is in fact a neural circuit, one part of which is only active during an allocation period; it is this part (the gate node) of each binding unit that projects inhibition to the blaster.

⁷ Although, there is evidence that the unavailability of the TAE may not be absolute. In particular, placing a cue before the T2 can attenuate the blink, suggesting that the cue fires the blaster even while T1 tokenization is ongoing (Nieuwenstein, Chun, van der Lubbe, & Hooge, 2005).

Each binding gate node also excites its corresponding token gate, which serves to activate the token. Continued activation of the type and token nodes is not necessary to maintain the binding once the binding trace node is active. However, if the type input from the TFL is too weak, the binding trace node will not reach its threshold and that item will not be encoded into WM.

4.7. Summary of the Neural-ST² Model

The following reviews the functional significance of each element of the model.

First Stage

1. *Input and masking layers.* Through inhibitory interactions, masking arises at the masking layer.
2. *Item layer.* Weak excitatory links between similar items facilitate priming.
3. *TFL.* This layer implements the salience filter. Task demand selects categories of items. The layer also sends feed-forward excitation to gate nodes of binding units and triggers the blaster.

Intermediary Stages

1. *Binding pool.* This provides a binding resource, which enables types to be bound to tokens. Binding units are implemented as gate–trace pairs, yielding a gated active memory. An inhibitory projection from binding gates to the blaster maintains the blaster offline during tokenization.

Second Stage

1. *Tokens.* These are also implemented as gate–trace pairs (called WM gate–traces), yielding another gated active memory.
2. *WM gate nodes.* These gate tokens access and are also the origin of (binding pool encoded) associations from tokens to TFL items. Strong lateral inhibition ensures that only one token is ever available.
3. *WM trace nodes.* These actively maintain tokens in a bound state. Thus, at encoding, they ensure that tokens can be bound only once per trial, and at recall, they denote that a token has been bound.

Blaster

On detection of a target, this mechanism causes a rapid and temporary increase in activation levels across the item layer and TFL.

4.8. Modeling the Empirical Landscape

The model reproduces a spectrum of AB data. We consider these results in turn.

The Basic Blink

Suppression of the blaster during tokenization is the mechanism that generates the blink. That is, the binding of T1 to a token can take long enough that a weak T2 has subsided in both the item layer and TFL before the blaster is available again. Targets at Lags 2 and 3 fall at the point of maximum impairment (see Figure 15a). However, the impairment decreases linearly through Lags 4, 5, and 6, as it becomes more likely that tokenization of the T1 finishes before T2 has decayed.

Lag 1 sparing (see Figure 15a) results from the T2 being close enough in time to T1 to take advantage of the (T1-initiated) blaster firing, allowing it to be encoded into the first token alongside T1. The loss in T1 accuracy at Lag 1 (see Figure 15c) arises because when T2 is very strong and T1 weak, binding can complete before T1 is strongly active, yielding a successful binding from Token 1 to T2 and no binding to T1. As Figure 15a shows, the results of the model largely conform to the conventional shape of the AB curve. Manipulating parameters can change the duration, depth, and onset of the blink, but the U shape is present whenever the model is operating in accord with its theoretical underpinnings.

Blanks

Placing a blank in the T1 + 1 slot attenuates the blink; see Figure 15a. An unmasked item generates a strong activation trace, a situation that greatly shortens tokenization. As a result, tokenization is more likely to have finished, which would release the blaster, before the T2 has decayed at the TFL. In contrast, T2 + 1 blank produces strong T2 traces that are capable of outliving the blink. Human data on T2 + 1 blanks do not exist, but we were able to compare the model with data from Giesbrecht and Di Lollo (1998), who examined the effect of placing T2 at the end of the stream; see Figure 15a. T1 + 2 and T2 + 2 blanks, on the other hand, have minimal effects on the trace strength of targets and therefore have little affect on the blink; see Figure 16.

Swaps

In accordance with human data, at Lag 1, the model is inaccurate at determining the order of the two targets; see Figure 15c. At Lag 1, typically, both T1 and T2 are bound to the first token and, in fact, to a small extent also to the second token. However, T2 is more often uniquely bound to the second token than is T1. It is this binding profile that allows the recall process to successfully disambiguate the order of presentation of the two targets on at least some of the Lag 1 trials, explaining why swaps are better than chance.

Missed T2 Processed Extensively

We adopt a compromise position on the fate of blinked T2s. None of the type layers in neural-ST² (masking, item, or TFL) are strongly suppressed during the blink. As a result, the model is capable of exhibiting priming between semantically related items even during the blink. That is, types are always, at least to some extent, extracted from a target, regardless of whether it is blinked. However, there are notable differences in the traces of seen and unseen T2s. A T2 that is seen during the blink has the virtue of

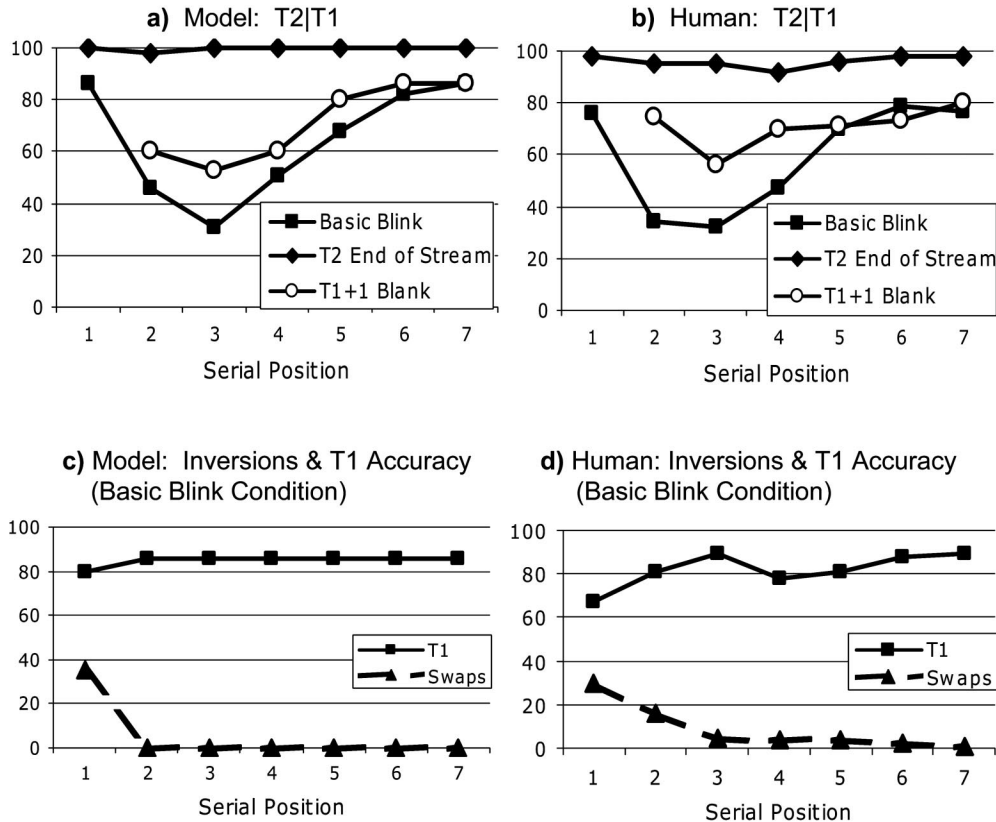


Figure 15. The model's performance (a and c) compared with human data (b and d). Target 2 (T2) performance (a and b) represents the accuracy in reporting T2 on trials in which Target 1 (T1) was reported. In c and d, the lines at the top of the graph show T1 accuracy, and the lines at the bottom denote the percentage chance for the reported order of T1 and T2 to be inverted. Horizontal axes represent lag, and vertical axes denote accuracy. In the T1 + 1 blank condition, there is no Lag 1 case, because that slot is blank. Human data are from "A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation," by M. M. Chun and M. C. Potter, 1995, *Journal of Experimental Psychology: Human Perception and Performance*, 21, p. 112 and 114. Copyright 1995 by the American Psychological Association. Adapted with permission of the author. In addition, the T2 end of stream data are from "Beyond the Attentional Blink: Visual Masking by Object Substitution," by B. Giesbrecht and V. Di Lollo, 1998, *Journal of Experimental Psychology: Human Perception and Performance*, 24, p. 1457. Copyright 1998 by the American Psychological Association. Adapted with permission of the author.

having a stronger trace than a missed T2, either by variation of input strength or by being excited by the blaster; see Figure 17.

The difference in activation strength, between missed and seen T2s increases as we progress through the Stage 1 layers. This is apparent in Figure 17, in which the difference between masking-layer missed and seen traces is very small but is slightly increased at the item layer and then further accentuated at the TFL. This is broadly consistent with Luck et al. (1996) and Vogel et al. (1998), who found no discernible difference between early ERP components (N1 and P1) for Lag 3 versus Lag 7 T2s. Such early ERP components would arise from ST²'s masking layer activations. They also found little discernible difference between the N400s elicited by Lag 3 and Lag 7 T2s, which would emerge from ST²'s item layer. However, Vogel et al. (1998) did observe a marked impairment in the P3s generated by Lag 3 versus Lag 7 T2s, which is likely to correspond to a combination of our TFL and token activations.

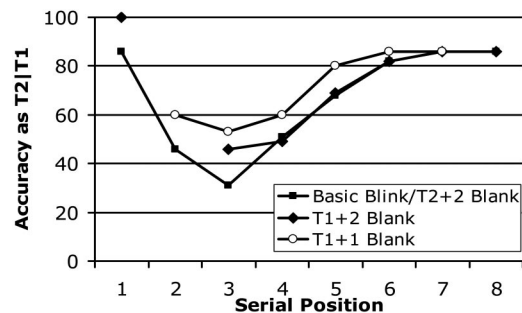


Figure 16. Model serial position curves for Target 1 (T1) + 2 and Target 2 (T2) + 2 blank compared with the basic and T1 + 1 blank conditions. The basic blink and T2 + 2 blank conditions are identical. Heightened T1 + 2 blank accuracy at Lags 1 and 3 arises because at Lag 1, the T2 is free from backward masking, whereas at Lag 3, it is free from forward masking.

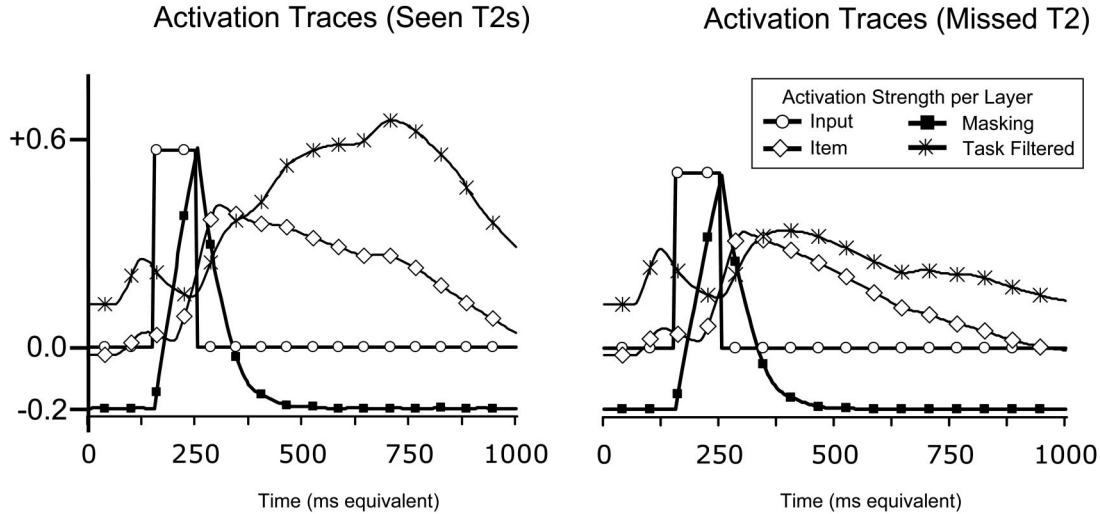


Figure 17. Activation profiles for first stage layers for seen versus missed Target 2s (T2s) presented at Lag 3 (i.e., during the blink). Profiles are averaged across all Target 1 (T1) times T2 values. The blips in item and task filtered layer activations are caused by the blaster firing (the first time of which is for the T1, before T2 has been presented). It is the small difference in bottom-up trace strength (evident in the input profiles) that generates the different profiles (and encoding outcomes) at later layers.

The only slight inconsistency with Vogel et al.'s (1998) results is that at our item layer, the difference in activation between seen and missed T2s is perhaps somewhat larger than the ERP N400 finding. As supported by Rolke, Heil, Streb, and Hennighausen (2001), we suggest that although missed T2s prime T3s, their priming will not be as strong as that for seen T2s, perhaps particularly in respect of the duration of the priming (Rolke et al., 2001). Furthermore, the comparison to Vogel et al.'s ERP profiles is somewhat muddled by two issues. First, the precise way to relate ERPs and neural network activation (especially activations generated by rate-coded models) is not well understood. Indeed, there is uncertainty about what exactly, neurobiologically speaking, is being recorded in ERP, especially from later ERP components (such as the N400 and P3; Fell et al., 2004). Second, Vogel et al. (1998) compared Lag 3 with Lag 7 N400s, whereas in Figure 17, we are comparing missed versus seen T2 traces at Lag 3, which, in a sense, is a purer measure of the effect of the blink. This is because the Lag 3 ERPs will contain a proportion of seen T2 trials (because T2 accuracy is never zero) and the Lag 7 ERPs will contain missed T2 trials (because T2 accuracy is never 100%). Thus, it may be more difficult to observe N400 differences between missed and seen T2s (which is the point of interest) using Vogel et al.'s (1998) lag-based differentiation.

Enhanced Processing of the T1 + 1 Slot

The blaster enhances the activation of the T1 and T1 + 1 slots; see Section 4.5. Accordingly, the blink is attenuated by the insertion, at the T1 + 1 position, of a semantic prime for the T2 (Chua et al., 2001); see Figure 18. This arises because the T1 + 1 distractor weakly excites its semantic associates (one of which is T2) via the lateral excitatory links in the item layer. We do not provide a quantitative fit to Chua et al.'s (2001) data because they

were working with a color marked task (and, in fact, identity rather than semantic priming), which is beyond the scope of neural-ST².

For comparison purposes, Figure 18 includes the performance of the model on the basic blink and T1 + 1 blank conditions. This makes clear that the priming benefit arising from a T1 + 1 distractor is relatively short-lived. Thus, if the (primed) T2 arrives at Lags 3 or 4, it benefits from its proximity to the semantically related T1 + 1 distractor. However, at later lags (e.g., 5 or 6), item layer activation of the T1 + 1 distractor has decayed sufficiently that the (primed) T2 receives little, if any, benefit.

4.9. Neurophysiological Correlates

There are three functional elements that we discuss in relation to proposed brain regions: Stage 1 (type processing), Stage

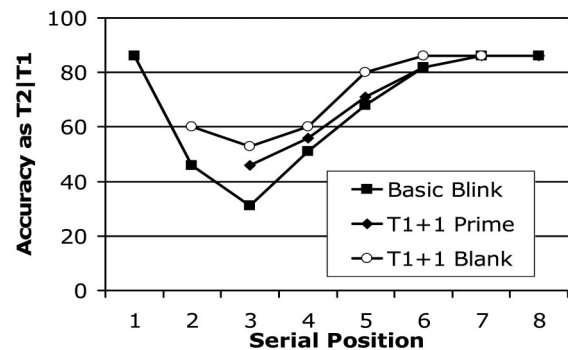


Figure 18. Model performance when the Target 1 (T1) + 1 distractor primes Target 2 (T2). We also include the model's basic blink and T1 + 1 blank simulations. We do not include Lags 1 or 2 for the T1 + 1 prime condition. Lag 1 is not a real data point, because the prime occupies this position, while Chua et al. (2001) did not consider Lag 2, because the proximity of the prime to T2 is a confound.

2 (token and WM maintenance), and the blaster (transient attention). First, there is a strong correspondence between the levels of ST²'s Stage 1 and sequential steps of the ventral visual processing pathway, which make up a cascaded highly parallel system (see Rousselet, Thorpe, & Fabre-Thorpe, 2004, for a review). In the visual system of the macaque, early stages (V1–V2) are thought to have small receptive fields, making them susceptible to spatially overlapped masking, much like the masking layer within ST². As one ascends the hierarchy of visual areas, to V4 and infero-temporal cortex, spatial specificity decreases as the processing gives way to progressively more identity-specific information (Rousselet et al., 2004). The same shift occurs within our model in progressing from the masking layer, with its strong spatial masking, to the item layer, which represents the semantic relationships between items.

Type nodes in the TFL correspond well to cells in the late stages of the ventral visual processing stream, a position supported by recordings from the inferior temporal cortex in macaques performing a WM task (Miller, Erickson, & Desimone, 1996). In these data, inferior temporal neurons selective for stimuli encoded into WM did not sustain activity during the presentation of other items. This is exactly the profile of activity produced by type nodes in ST², as activation of a type is only necessary during encoding and retrieval.

Sustained activity during maintenance is the province of tokens and the binding pool. In the prefrontal cortex, Miller et al. (1996) found cells that have sustained activity and are insensitive to stimulus identity, as would arise from trace nodes of tokens. In contrast, other cells in the prefrontal cortex sustained activity and were stimulus selective. These cells could correspond to trace nodes within the binding pool.

Neurophysiological support for trace nodes in humans can be seen in a recent functional magnetic resonance imaging study of the AB (Marois, Do-Joon, & Chun, 2004). Of particular note, Marois et al. (2004) identified a notably greater response in a lateral frontal region for seen versus missed T2s, which remained pronounced even at the end of the recording period. This may be the marker of sustained activation of two active tokens (in the seen case) versus one (in the missed case). Kranczioch, Debener, Schwarzbach, Goebel, and Engel (2005) made a similar point. They demonstrated that in an AB setting, when participants saw the T2, activity in frontal and parietal areas (inferior frontal gyrus, lateral frontal cortex, and inferior parietal lobe) was elevated in a sustained manner, whereas there was no sustained elevation in the ventral stream (lateral occipital complex and fusiform gyrus). In our model, WM trace nodes would perform such sustained activation. In contrast, areas of the brain that would correspond more closely to type processing (parahippocampal place area in Marois et al., 2004, and fusiform gyrus–lateral occipital complex in Kranczioch et al., 2005) showed temporary differences between T2 seen and T2 missed trials. The localization of the token system to prefrontal areas is also consistent with an extensive literature of imaging and lesion studies that demonstrate how prefrontal areas contribute to WM encoding, maintenance, and retrieval (Desimone, 1996; Passingham & Saka, 2004). In addition, Petrides (1995) demonstrated that monkeys with lesions of the middorsal part of the lateral frontal cortex have a permanent deficit in their ability to recall the temporal order of items.

The neurophysiological implementation of transient attention may correspond to a putative attentional mechanism comprising right-lateralized areas of the temporo-parietal junction (TPJ) and ventral frontal cortex (VFC); for a review, see Corbetta and Shulman (2002). These areas of the brain are thought to detect both exogenously and endogenously salient targets and rapidly orient attention to them. For example, in a recent imaging study by Serences et al. (2005), participants monitored a central RSVP stream for red letters while ignoring two flanking streams. When flanking distractors switched from gray to a nonred color, they stood out strongly but failed to capture attention. Only when the flanking items were the same color as the target was a contingent capture of attention observable as a reduction in target detection. On presentation of red (target colored) flanking distractors, the right TPJ and VFC were strongly activated, in contrast to when nontarget colored distractors were presented. This cortical circuitry may work in concert with subcortical mechanisms, such as the locus coeruleus (LC; see Section 6.8). As described by Aston-Jones, Rajkowski, and Cohen (2000), within a visual discrimination paradigm, monkey LC neurons are activated with a temporal profile that matches closely what our blaster would predict. Namely, in such a paradigm, LC neurons were strongly activated for a period of 100 ms, peaking about 150 ms after a target was presented. Following this activation, LC neurons appeared to be inhibited for a further 100 ms. These data suggest that the LC and the TPJ–VFC networks may cooperate to rapidly deploy transient attention in response to detected targets with cortical mechanisms filtering stimuli through the frontally mediated task set and triggering the LC to coordinate the deployment of attention.

VALIDATION OF THE MODEL

5. Predictions and Their Empirical Validation

Section 4.8 showed that neural-ST² does a good job of modeling existing AB findings. Here we generate predictions from ST² and validate these predictions empirically.

5.1. Prediction 1

Within the ST² theory, the temporal dynamics of AB sparing are dictated by the TAE, which fires for a (broadly) fixed window of time. The most direct prediction of this mechanism is that sparing is primarily dictated by the temporal gap between targets, rather than their sequential separation. By doubling the rate of presentation from 100 to 50 ms per item, we should observe that sparing extends out to Lag 2. That is, items are spared at short lags not because they are directly adjacent to the T1 but because they occur within 100 ms of it.

Figure 19 (on the left) demonstrates the model's performance at 100- and 50-ms SOAs. These data are sorted by the time between targets (i.e., target onset asynchrony [TOA]), rather than by the number of items between targets (i.e., lag). The 100-ms/item data

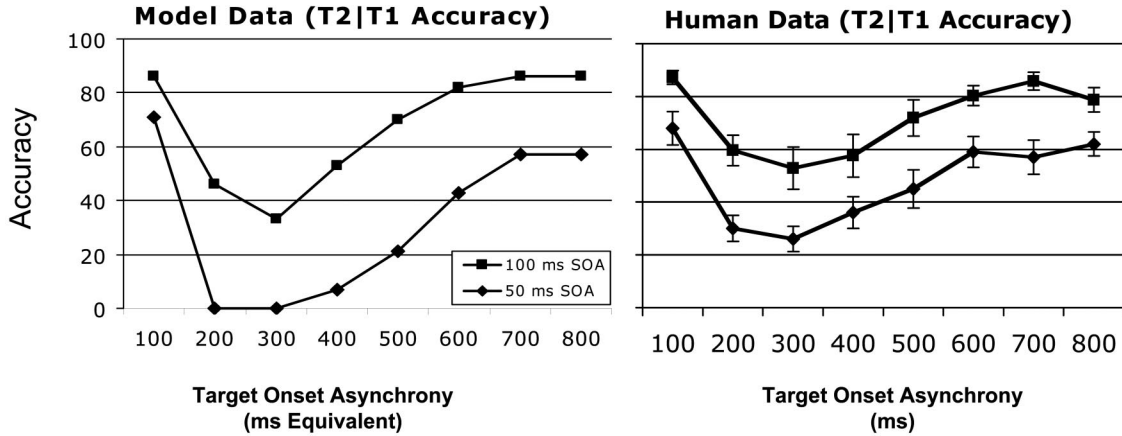


Figure 19. Data from the neural simultaneous type, serial token model and humans showing blink curves when items are presented (in the usual manner) at stimulus onset asynchrony (SOA) of 100 ms and when the presentation rate is doubled to an SOA of 50 ms.

are the basic blink condition of Figure 15. For the faster presentation rate, data are shown for TOAs of 100–800 ms, corresponding to Lags 2, 4, 6, 8, 10, 12, 14, and 16.⁸ Note that there is strong sparing even for Lag 2, and the general time course of the blink is similar to that for the slower presentation rate. The reduced baseline accuracy (easily seen at Lag 8) is the result of the weaker traces produced by rapid presentation rates.

To test this prediction, we ran an experiment, described in the 54-ms SOA experiment in Section A.3 of the Appendix, presenting participants with letter targets in a stream of digit distractors at 54 ms/item. As predicted, the AB conformed to the conventional temporal characteristics, with sparing at a 100-ms TOA and recovery well under way by a 500-ms TOA; see Figure 19 (on the right). These data are directly compared with the results of a second experiment (see the 94-ms SOA experiment in Section A.3 of the Appendix) that used a 94-ms SOA. Comparing these results with a 2×8 analysis of variance revealed a main effect for lag, $F(7, 176) = 12.34, p < .00001$, and for presentation speed, $F(1, 176) = 76.69, p < .00001$, but no interaction, $F(7, 176) = 0.39, p > .90$. The lack of an interaction indicates the absence of an effect of presentation speed on the time course of the AB. These results stand counter to models of the AB that place primary importance on the T1 + 1 distractor as an initiator of the blink, including the interference theory (Shapiro, Arnell, & Raymond, 1997) and the temporary loss of control theory (Di Lollo, Kawahara, Shahab Ghorashi, & Enns, 2005).

5.2. Prediction 2

The onset of transient attention is delayed relative to T1 and extends for some time thereafter. Typically, this time course ensures that a T2 at Lag 1 is perceived more accurately than T2s that themselves trigger the blaster (i.e., at Lag 8). In effect, T1 acts as a cue for T2 in the same way that cues in Nakayama and Mackeben (1989) improved performance for the cued item. We observe this pattern within the model. If we switch to a nonconditional analysis (T1 may fire the blaster even when it is too weak to be reported), we see an improvement in T2 at Lag 1 relative to all other lags. In

the model, unconditional T2 accuracy stabilizes at 86% for Lags 7 and 8 but is 89% at Lag 1. This effect is even more pronounced for a faster presentation rate as performance is farther from ceiling. In the model, the baseline accuracy of 58% at Lags 14 and 16 increases to 75% at Lag 2 (100-ms TOA) for a 50-ms SOA.

This prediction is supported by a reanalysis of the data collected for Prediction 1 (methods reported in Section A.3 of the Appendix), which examined the raw (i.e., nonconditional) accuracy of T2. Using a planned series of t tests for the slow condition, we found a marginal, but significant, improvement of T2 during sparing. T2 was 90% ($SE = 2\%$) at Lag 1 and 83% ($SE = 3\%$) at Lag 8. This difference is significant, $t(13) = 2.74, p < .02$.

Turning to the fast presentation rate, in which we expect an even more pronounced effect, T2 at Lag 2 (equivalent to Lag 1 at the slower rate) was 75% ($SE = 5\%$), and at Lag 16 (equivalent of Lag 8) was 59% ($SE = 4\%$). This difference is significant, $t(9) = 3.06, p < .015$. These results confirm that not only is T2 spared at Lag 1, but it can also exhibit a net accuracy improvement. The results from Predictions 1 and 2 support the TAE hypothesis. Indeed, it is hard to see how raw T2 could be superior at early lags to postblink recovery without a T1-initiated transient enhancement.

5.3. Prediction 3

As discussed in the *Lag 1 Sparing* subsection of Section 3.3, we claim that sparing may be indicative of a failure of selective attention. That is, the processing mechanisms revealed by the AB may exist to exclude T2 when it threatens to interfere with T1, but at Lag 1, there is insufficient temporal resolution to exclude the second target. Putative failure of attention appears as sparing at Lag 1 for two reasons. First, in the majority of AB studies, either temporal order of T1 and T2 was implicit in the task switch

⁸ To pull the 50-ms model data above floor, we have increased the input strength compared with the 100-ms SOA conditions. This is not surprising because it is consistent with the existence of a nonlinear relationship between bottom-up trace strength and length of item presentation/SOA.

between the targets (e.g., Raymond et al., 1992), or order report was not requested of participants (e.g., Chun & Potter, 1995). Second, targets in AB studies have almost always been unique, individually discriminable items that cannot be miscombined into other possibly valid targets. What this means is that T1 and T2, even when bound into the same token at Lag 1, could be (unambiguously) separated in a post hoc retrieval.

We predict that an AB experiment that requires participants to maintain both the temporal order of T1 and T2, as well as the binding of subelements within targets, will produce accuracy at Lag 1 that is worse than at any other lag. The experiment to test this prediction demands that participants report the identity of both T1 and T2, in the correct order. Furthermore, T1 and T2 each consist of a letter pair, which must be accurately paired at recall. Thus, given a T1 of *AB* and a T2 of *CD*, responses of *CD AB* would be counted as errors, as would *AD CB*. We ran this experiment, as described in the letter pairs experiment in Section A.3 of the Appendix.

Figure 20 shows that at Lag 1, T2 exhibited no appreciable sparing and T1 performance was at just 20%, having dropped from a baseline of about 65%. By combining these [the $(T1 + T2)/2$ curve], we indexed the efficiency with which participants extract correctly bound targets (T1 or T2) in the proper temporal order. There was a sharp reduction in $(T1 + T2)/2$ at Lag 1 that was more severe than the decline at any other lag; see Figure 20. In a planned t test, Lag 1 was significantly lower than Lag 2, $t(11) = 7.5$, $p < .00001$.

These data strongly support our position that use of the term *sparing* to describe the Lag 1 case is, in many respects, an artifact of the form of classic AB experiments. We mean this in the sense that in these experiments, (a) the order in which targets were detected was not assessed and (b) component features of T1 and T2 were not confusable. Our results show that when these constraints are removed, sparing disappears at Lag 1; indeed it is the lag at which the least information is extracted. We thus have obtained evidence to support our position that the blink is a mechanism to prevent the arrival of T2 from interfering with the ongoing T1 binding. Furthermore, Lag 1 is a breakdown of the system in the sense that the T2 arrives so rapidly after T1 that the system is unable to withhold attentional resources from T2.

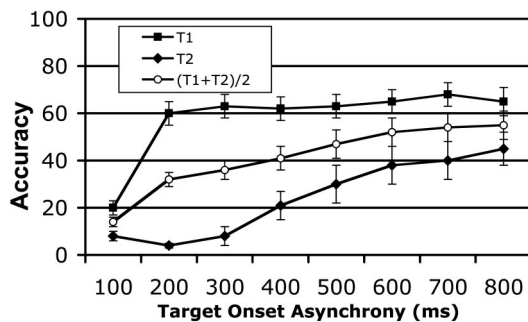


Figure 20. Human serial position curves for the letter pairs task. Raw (i.e., nonconditional) Target 1 (T1) and Target 2 (T2) accuracy is shown, along with a measure of overall information extraction, calculated as the mean of T1 and T2 accuracy.

COMPARISON WITH OTHER MODELS

There are a number of AB models, which we compare with ST² in this section. We begin by considering the applicability of three general neural mechanisms. Then we consider five formal models, and finally, we discuss four informal theories of the AB.

6. Neural Models

General Neural Mechanisms

We consider the applicability of three neural mechanisms for modeling the AB.

6.1. Active WM

WM is often modeled as sustained neural activation (Davelaar et al., 2005; Davelaar & Usher, 2004). In a localist approach, this is realized by a self-sustaining node; see Figure 7a. Our TFL uses aspects of localist active memories, such as self-loops. However, we also include off nodes with relatively rapid activation dynamics (certainly in relation to the decay rate of WM). As a result, TFL nodes self-sustain only for brief periods (a few hundred milliseconds) before being suppressed by their off nodes. So, why not do away with these off nodes and implement an active WM at our TFL?

Unfortunately, pure active memories do not reflect the types–tokens distinction. For example, because type representations are not freed up, pure active memories have difficulties representing multiple instances of the same item; that is, they are permanently repetition blind. In addition, there is no mechanism available to represent temporal order. (These issues were discussed in Section 2.3.) Extra mechanisms could no doubt be added in response to these limitations. However, such an approach would not be obviously simpler or more theoretically appealing than the tokenization approach we are advocating.

6.2. Winner Take All

A simple way to generate an AB would be to enforce lateral inhibition between task-relevant items at, what would become, a winner-take-all layer. For example, in our model, we could impose strong lateral inhibition at the TFL and view that as the output from the model; that is, we could remove the WM token system and simplify TFL nodes by removing their self-loops and shut-off systems. At such a winner-take-all layer, T1 activation would suppress a T2 arriving during the blink. In order to explain Lag 1 sparing, this approach would have to somehow argue for a delayed onset of T1 inhibitory pressure. In addition, the approach would explain blink recovery through T1 activation decay, which would eventually release the layer from inhibition. Although pleasingly simple, this approach is unsatisfactory for a number of reasons.

WM encoding. Such an approach would not explain how items are encoded (and maintained) in WM. Rather it would just generate short-lived activation traces at the (final) winner-take-all layer. One would then have to argue that activation that crosses a particular threshold at the winner-take-all layer would be interpreted as reported. But, there would be no mechanism provided to explain how those items would be maintained until the point at which report occurs. In particular, by the nature of this approach, item nodes cannot self-sustain at the winner-take-all layer, because

if they did, there would be no blink recovery, because T2s would be permanently suppressed by T1 activation.

Late AB bottleneck. In addition, it is unclear how to reconcile such an approach with the known late locus of the AB; see Section 3.2. It is difficult to be unequivocal about this point because it is dependent on the interpretation imposed on layers of the model. However, if the winner-take-all layer was interpreted as a component of the representation of a type, then a missed T2 would have a greatly reduced (compared with seen T2s), and probably non-existent, type activation. This may well imply absence of a semantic representation of missed T2s.

Types–tokens distinction. A winner-take-all approach would also fail to code temporal order. Although, the approach could potentially avoid the problem possessed by active memories of permanent repetition blindness. This is because target activation would eventually decay (this is required to ensure blink recovery), and thus, a (sufficiently delayed) repetition could be represented. However, avoidance of permanent repetition blindness is only obtained because a temporally sustained representation is not generated; see *WM encoding* above.

Lag 1 sparing. Finally, a winner-take-all approach would not give an illuminating explanation of Lag 1 sparing. That is, if one views the lateral inhibition as reflecting interference between T1 and T2, then one is still left with the problem of explaining why T1 and T2 do not interfere at Lag 1.

6.3. TAE Unavailability as a Refractory Period

A number of models use approaches similar to our TAE (e.g., see Section 6.8)—a generalized (in terms of types) and short-lived enhancement that is initiated by the detection of an AB target. Furthermore, they all generate a blink by the TAE being offline when the T2 arrives. There is though an important issue concerning how the TAE is held offline. In ST², two mechanisms contribute to the TAE being held offline: (a) a blaster refractory period and (b) ongoing tokenization. The first of these is common to all the TAE based blink models. However, it is only ST² that also includes the second. This is important because refractory periods have a fixed time course, which is locally controlled and not tied to ongoing processes elsewhere in the model.

However, tying TAE unavailability to tokenization fits with the position that the system tries to prevent a second target from interfering with ongoing tokenization. This, in turn, fits with the ST² principle of ascribing discrete episodic contexts. In more practical terms, ongoing tokenization causing the TAE to be withheld explains why blink length or depth varies with bottom-up trace strength. That is, the shorter the time to tokenize, the shorter TAE unavailability, and thus, the shorter the blink. Approaches that just tie TAE unavailability to its refractory period lack this link between tokenization and blink length. As a result, they typically find it more difficult to explain how unmasking T1 attenuates the blink.

Formal Neural Models

We now discuss the available formal neural models of the AB.

6.4. The Gated Auto-Associator

The model. Chartier, Cousineau, and Charbonneau (2004) presented a connectionist model of a color marked digits AB task; see

Figure 21. It contains an input layer, which feeds into two competitive networks, performing number and color identification, respectively. The number identification system feeds into an auto-associator, which acts as a WM. When the representation of a digit becomes active in this auto-associator, a learning rule is used to adjust the weights to auto-encode (into WM) the activation pattern.

Gating mechanism. A gating mechanism sits between the number identification system and the auto-associator. The opening of this gate is regulated by a comparison system, which compares the color being identified with the target color. If the colors match, the gate is opened, and the digit currently active at the identification network is encoded into WM. However, opening the gate starts an inhibitory process, which ensures that the comparison generated by the T1 + 1 item takes significantly longer than for the T1. Comparison efficiency progressively improves from the T1 + 2 item onward until it eventually returns to its original level.

Auto-association. Items are maintained in WM via auto-associator weights. The strength of these weights, that is, the strength of encoding, is a function of the extent to which the gate is open while the item is represented at the identification layer. The system recalls the two items that are most strongly encoded into WM, that is, that possess the strongest weights.

How the blink arises. The key mechanism that generates the blink is the inhibitory process. Thus, the increase in comparison time resulting from the inhibitory process ensures that the gate opens later for a T2 presented during the blink. Such T2s are thus only weakly encoded into WM, and it is likely that a distractor is more strongly encoded, which would be recalled in preference to the T2.

Blink recovery is governed by the length of time it takes this inhibitory process to subside. Lag 1 sparing is obtained because a T2 at Lag 1 benefits from the gate remaining open following T1 encoding. Thus, in this case, T2 encoding is largely independent of the T2 comparison process.

Assessment and comparison to ST². Chartier et al. (2004) provided a valuable AB model. However, it has a somewhat different intent from neural-ST², because it is targeted at color marked data, rather than at letters-in-digits data. Thus, it is difficult

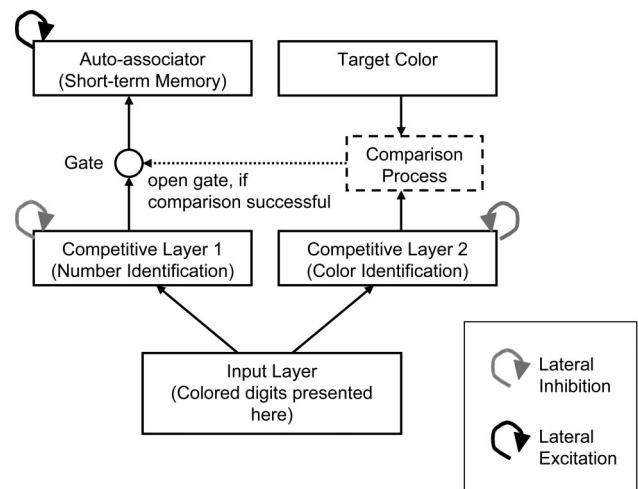


Figure 21. The gated auto-associator model.

to make a direct comparison, but we offer the following observations.

First, the approach only reproduces a small set of AB phenomena when compared with neural-ST². It reproduces a blink curve with Lag 1 sparing.

Although not a strictly active memory approach, because items are stored as weights rather than as sustained activation, the gated auto-associator does suffer some of the problems of such approaches. For example, it seems that the model would also be permanently repetition blind. This is because both instances of an item would generate identical auto-associator activation. Thus, a repeated item would be extremely strongly encoded (in the sense of weight strength), but there would be no representation of multiple tokens.

In addition, it seems that the gated auto-associator cannot represent serial order. Although, there is an indication to the contrary in Chartier et al. (2004), in which they claimed that their model generates inversion errors (i.e., swaps); however, no supporting mechanism or data are provided.

The gated auto-associator relies on a very fast weight change process, which would be somewhere in the 100-ms time frame. This is difficult to support from a neurobiological standpoint; see Section 2.5.

The gated auto-associator does not contain an analogue of our masking system. Thus, it is difficult to see how the model could reproduce blink attenuation with target + 1 blank.

Finally, the approach seems to be consistent with a late bottleneck view of the AB. Specifically, assuming that the competitive layers perform all processing prior to WM encoding, then missed T2s should obtain identical preencoding activation profiles as reported T2s.

6.5. The Global Workspace Model

The model. The Global Workspace model provides an elegant and compelling theory of conscious perception and attentional control (Dehaene, Sergent, & Changeux, 2003). The model includes neural processing pathways (see Figure 22) from early sensory regions (Areas A and B), for example, the visual cortex, through to higher association areas of the temporal, parietal, frontal, and cingulate cortex (Areas C and D; Dehaene et al., 2003). Stimuli compete to recruit a global workspace sitting at the top of these pathways that, once recruited, affords global exclusive access. That is, the workspace “mobilises excitatory neurons with long distance axons, capable of interconnecting sensory and high-level areas into global brain-scale states” (Dehaene et al., 2003, p. 8520).

Exclusivity of brain-scale states (i.e., that consciousness is unitary) arises because neurons in the workspace inhibit surrounding neurons (the lateral inhibition in Figure 22). Furthermore, reverberating activation between constituent neurons generates a pressure for brain-scale states to be self-sustaining. The model is biologically detailed; for example, it uses spiking neurons and is anatomically prescribed. In particular, cortico-thalamic columns act as building blocks for the model; see Figure 22.

How the model blinks. The AB version of the global workspace model contains processing pathways from visual input for both T1 and T2; see Figure 22. The firing rate of pyramidal cells coding T2 in the highest two areas (C and D in Figure 22) is taken as an index of T2 reportability. The blink arises because the

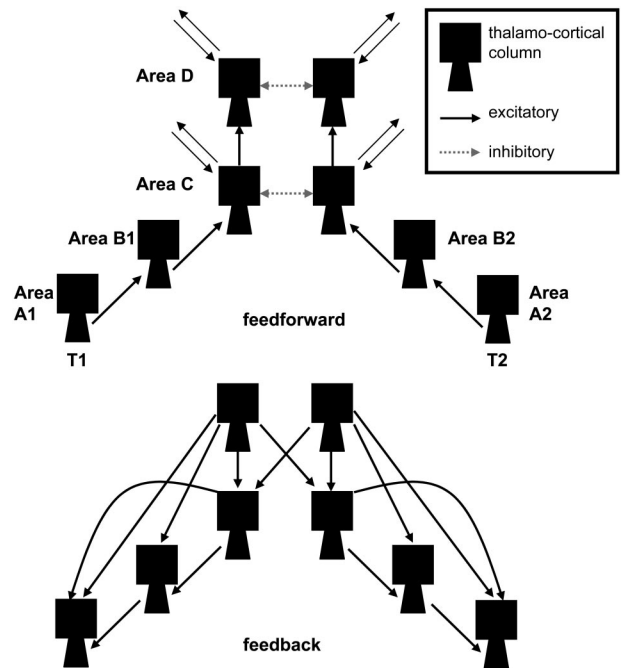


Figure 22. Processing pathways in the global workspace model. This diagram is from “A Neuronal Network Model Linking Subjective Reports and Objective Physiological Data During Conscious Perception,” by S. Dehaene, C. Sergent, and J. P. Changeux, 2003, *Proceedings of the National Academy of Sciences, USA*, 100, p. 8521, Figure 1B. Copyright 2003 by the National Academy of Sciences, USA. Adapted with permission.

brain-scale state generated by the T1 suppresses competing stimuli and, thus, reduces their reportability index. Recovery from the blink arises when the T1 brain-scale state has subsided, freeing the workspace for late-lag T2s.

Assessment and comparison with ST². We now discuss the contribution of the global workspace model in the context of the AB. The model does not exhibit as broad a spectrum of data as ST²; it generates a blink curve, and a fit with electrophysiological data is highlighted. We return to both these issues in later points.

Dehaene et al. (2003) generated a U-shaped blink curve. However, the blink curve in Figure 3A of Dehaene et al. (2003) is rather short, with accuracy returning to baseline by a 200-ms T1–T2 lag, whereas human blink curves tend to return to baseline by Lag 5, that is, a T1–T2 lag of around 500 ms (e.g., see Figure 15).

Figure 3A and the results section of Dehaene et al. (2003) indicate that they have obtained Lag 0 sparing; that is, they found that T2 is spared if it is presented simultaneously with T1, which is not a typical AB data point in which items are presented in sequence. In addition, the AB curve in Figure 3A is at its deepest point at T1–T2 lags of 50 ms and 100 ms, which is the region at which Lag 1 sparing is obtained in human data; see, for example, Figure 15. However, the parameter settings used have been drawn from the monkey literature, and as indicated by the authors, they may not be consistent with the human system.

Dehaene et al. (2003) acknowledged the limitations that arise from not implementing masking. This means that this model cannot address the impact of blanks in the stream. In addition,

because RSVP distractors are not represented at all in the global workspace simulation, the model is unable to address the capacity of $T1 + 1$ distractors to prime future targets (Chua et al., 2001).

As previously highlighted, the key mechanism in the global workspace model that enables it to blink is the lateral inhibition between thalamo-cortical columns in Areas C and D. Effectively, this means that the model implements a winner-take-all solution. Thus, the model has many of the characteristics discussed in Section 6.2. Finally, there is no representation of serial order in the model.

6.6. The Conflict Monitoring Model

Elements of the model. Battye's (2003) conflict monitoring model uses a stimulus-response pathway, which includes localist input and output layers connected in a one-to-one fashion (see Figure 23). Attentional control arises from an attention layer, which contains two nodes, denoted *location* and *letter*. The former of these gives an attentional boost to stimuli presented in a particular spatial location, whereas the latter foregrounds the task set. This foregrounding is realized through projections from each attention node to the relevant input nodes. In fact, because all items are presented in the same spatial location in RSVP, the location node projects to all input nodes. Note that the attention to input layer projections are bidirectional. Thus, in addition to top-down (endogenous) attentional control, the model also supports bottom-up (exogenous) control (from the input to the attention layer). This bidirectionality generates a transient attentional response (TAR), similar to ST^2 's TAE.

The final mechanism implemented is conflict monitoring, which takes inspiration from theories of cognitive control (Botvinick et al., 2001). Conflict is a measure of the extent to which multiple responses are coactive and are thus interfering. Note that the strength of lateral inhibition at the output layer is adjusted according to conflict. Thus, when there is response conflict, competitive pressures are increased, so that the most active response can more easily dominate.

How the model blinks. There are two mechanisms (discussed below) that cause the blink:

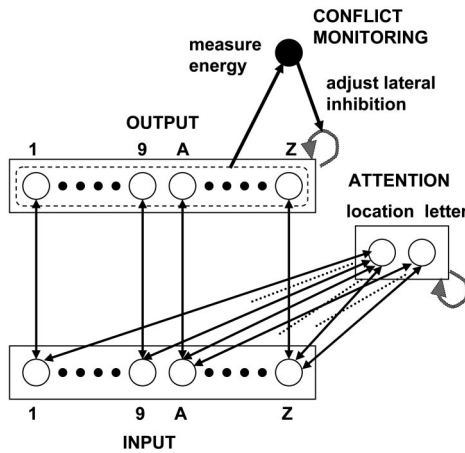


Figure 23. Battye's (2003) conflict monitoring model. Gray self-loops indicate a layer with lateral inhibition.

1. Adjustment of output layer lateral inhibition because of conflict.
2. The TAR.

Lateral inhibition and conflict is the first mechanism. As previously highlighted, the value of output layer lateral inhibition is proportional to the current value of conflict. Thus, coactivation, at the output layer, of $T1$ and the $T1 + 1$ distractor generates a short period of high conflict, which pushes output layer lateral inhibition up, closing the door on the $T2$. The door stays closed while the $T1$ and $T1 + 1$ distractor responses are coactive.

The TAR is the second mechanism. In the Battye (2003) model, the $T1 + 1$ distractor is only strong enough to cause conflict at the output layer because it benefits from the $T1$ initiated TAR. When $T1$ is active at the input layer, it excites the location node in the attention layer; but because this node projects to the whole of the input layer, a window of recurrent activity boosts the $T1 + 1$ distractor. However, once it has fired, the TAR enters a refractory period, which contributes to the blink.

Thus, in Battye's (2003) model, the blink arises from the following sequence of events. (a) The $T1$ initiates a brief window in which the input layer is boosted (the TAR); (b) the $T1 + 1$ item benefits from this boost; (c) coactivation of $T1$ and $T1 + 1$ at the output layer generates conflict; (d) this causes increased lateral inhibition at the output layer; (e) the now strong lateral inhibition causes $T2$ s to be suppressed by the already strongly active $T1$ and (to a lesser extent) $T1 + 1$; and thus, (f) the $T2$ yields only a weak output layer representation. Blink recovery arises because the $T1$ and $T1 + 1$ output layer representations eventually subside, which releases the system from conflict and resets lateral inhibition to zero. In respect of Lag 1 sparing, a $T2$ at Lag 1 will benefit from the TAR and will yield a strong output layer representation before conflict is registered.

Assessment and comparison to ST^2 . A strength of Battye's (2003) model is its simplicity and that it fits within an accepted connectionist-modeling framework: GRAIN (McClelland, 1993). A number of aspects of Battye's model have similarities to ST^2 mechanisms. Perhaps most significant, Battye's attention layer could be viewed as a conflation of our task demand and TAE. Specifically, Battye's letter node is an almost direct analogue of neural- ST^2 's task demand unit, whereas his location node is a TAE. This said, an important difference with ST^2 is that, effectively, Battye's TAR has a fixed refractory period. As a result, Battye's approach suffers the problems we discussed in Section 6.3 associated with not tying TAE unavailability to ongoing tokenization.

This difference is reflected in the manner in which Battye's (2003) model generates blink attenuation with $T1 + 1$ blank. The ST^2 model explanation is that increased bottom-up trace strength of an unmasked $T1$ speeds up $T1$ tokenization. In contrast, in Battye's model, a $T1$ followed by a blank will generate less conflict. This is because, in the $T1 + 1$ blank case, the distractor after $T1$ (at $T1 + 2$) obtains a reduced benefit from the TAR and, in any case, arrives a little too late to generate substantial conflict with the $T1$ at the output layer.

In addition, there is more of a serial (by-item) element to the blink in Battye's (2003) model. This arises because in Battye's model, the blink is initiated by the increase in conflict caused by

the arrival of the $T1 + 1$ item. Thus, a delay in the onset of this distractor, which arises in the $T1 + 1$ blank condition, will itself delay the onset of increased output layer conflict. Battye justified the approach on the grounds that in Chun and Potter (1995) there is evidence of Lag 2 sparing in the $T1 + 1$ blank condition. Although, it should be noted that the ST^2 model obtains a degree of Lag 2 sparing in the $T1 + 1$ blank condition (see Figure 15a) without subscribing to this theoretical position. Also, our experimental verification of Prediction 1 (see Section 5.1) stands against the serial aspect of Battye's model. Although Battye's conflict monitoring approach certainly makes an important contribution, there are a number of limitations.

1. Battye's (2003) model does not generate as broad a spectrum of data as the ST^2 model. In particular, there is no consideration of impaired $T1$ performance and increased swaps at Lag 1 and blink attenuation with $T2$ unmasked. However, inclusion of the TAR suggests that Battye's model is theoretically consistent with increased processing of the $T1 + 1$ distractor (Chua et al., 2001).
2. The theoretical justification for using a conflict monitoring approach in the context of the AB is not as clean as it may initially seem. In particular, Battye (2003) hypothesized a much faster adjustment of cognitive control than that previously proposed. For example, in Botvinick et al. (2001), cognitive control is updated per trial and not at the 100-ms conflict detection to adjustment of control speed that Battye required.
3. Probably most significant, Battye's (2003) approach seems incompatible with a late bottleneck. This is because in Battye's model the key to the blink is the enforcement of strong lateral inhibition in order to suppress the $T2$. Thus, the model effectively uses a winner-take-all strategy, and thus, it suffers the problems discussed in Section 6.2, which include incompatibility with a late bottleneck.
4. As was identified in Section 6.2 as a general problem of winner-take-all models, Battye's (2003) conflict monitoring model also does not provide an explicit model of how items are encoded, maintained, and retrieved from WM.

6.7. Corollary Discharge of Attention Movement (CODAM) Model

The model. Fragopanagos, Kockelkoren, and Taylor (2005) modeled the AB in the context of Taylor's (2002) CODAM model (Taylor & Rogers, 2002), which has made an important contribution to the modeling of attentional control. The model contains a bottom-up pathway, comprising the input and object map modules, along which input stimuli have to pass in order to reach the WM module; see Figure 24. Items are encoded into WM in an active memory fashion (see Section 6.1). Critically, items in the object map require an attentional boost to successfully progress to the WM. This boost is provided by an attentional control signal generated by the inverse model controller (IMC).

How the model blinks. In an AB setting, the attentional control signal is withheld for the $T2$, to prevent its encoding interfering

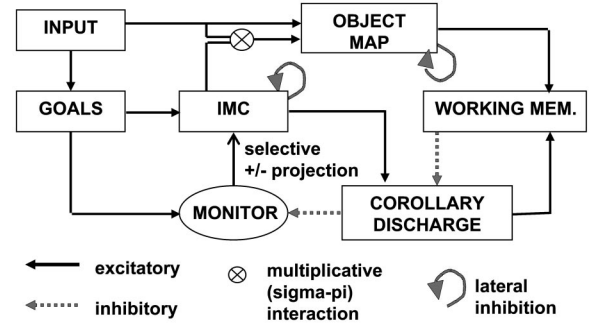


Figure 24. The corollary discharge of attention movement model used to simulate the attentional blink. The attentional control signal from the inverse model controller (IMC) to the object map is modulated by the input to object map projection. The monitor to IMC projection is selective—that is, it excites target items while inhibiting all other IMC items. This diagram is from “A Neurodynamic Model of the Attentional Blink,” by N. Fragopanagos, S. Kockelkoren, and J. G. Taylor, 2005, *Cognitive Brain Research*, 24, p. 570, Figure 1a. Copyright 2005 by Elsevier. Adapted with permission.

with the encoding of the $T1$. Understanding this mechanism requires us to consider the monitor, endogenous goals (part of the goals module), and corollary discharge; see Figure 24. The monitor has the capacity to suppress items in the IMC that are interfering with ongoing target encoding. The mechanism by which the monitor determines when it should selectively inhibit the IMC is slightly more involved, but the key aspect is that the monitor is continuously computing a comparison between the activity in the endogenous goals and the corollary discharge. The former of these maintains a representation of the current target, and the latter maintains a predictor of the attended stimulus by taking a copy of the attention control signal, that is, the IMC.

The blink arises because $T1$ remains the current target until its encoding at the WM layer is complete. Thus, if the system attempts to boost the $T2$, through the IMC, before $T1$ encoding is complete, the monitor will detect an error, because the current endogenous goal is $T1$ and the corollary discharge is representing $T2$. As a result, the monitor will suppress the $T2$ in the IMC and, thus, retract its attentional boost, which will in turn prevent $T2$ from being successfully encoded into WM. When $T1$ WM encoding is complete, the system recovers from the blink by handing the endogenous goal from $T1$ to $T2$. Consequently, the monitor will not register an error when comparing the current endogenous goal (now $T2$) with the corollary discharge representation of $T2$.

Assessment and comparison to ST^2 . There are some similarities between the CODAM approach and the ST^2 model. First, the activation profiles shown in Fragopanagos et al. (2005) suggest that the CODAM approach is consistent with a late bottleneck. Second, both hypothesize that this bottleneck arises to prevent interference during WM encoding. However, in instantiating this theoretical position, the two are rather different.

A strength of the CODAM approach is that it simulates the AB within the context of a broad scope and mature model of attentional control, which also has ties to neurobiology. In addition, it produces a larger spectrum of data than the other approaches discussed in this section. It generates a basic (U-shaped) blink curve with Lag 1 sparing as well as blink attenuation with $T1 + 1$

blank. However, the CODAM approach does not generate as broad a spectrum of data as the ST² model. In particular, it does not reproduce blink attenuation with T2 unmasked, decline in T1 performance at Lag 1, increased temporal order confusion at Lag 1, and increased processing of the T1 + 1 distractor.

In some senses, CODAM and the ST² model have different ambitions: The starting point for CODAM is neurobiology, whereas ST² began as a cognitive-level explanation of the AB and temporal attention in general. Thus, ST² has been driven by a desire to reflect, for example, the types-tokens distinction. In addition, Fragopanagos et al.'s (2005) approach is subject to the characteristics of pure active memory approaches (see Section 6.1): permanent repetition and temporal order blindness.

In the CODAM approach, recovery from the blink is tied to task handover: from the T1 to the T2 task. This may have some degree of plausibility in the context of blink paradigms with a task switch (e.g., Raymond et al., 1992). However, other AB paradigms, such as the letters-in-digits task, do not possess a task switch between the two targets, yet there is still a blink and a blink recovery. In addition, it has been argued that a task switch confounds the AB (Chun & Potter, 2000; Potter, Chun, Banks, & Muckenhoupt, 1998). Thus, as implemented, the CODAM approach has difficulty fully explaining blink recovery, certainly in the context of experimental paradigms without a task switch.

6.8. The LC Model

The model. This recently proposed model explains the AB according to the functioning of the locus coeruleus (LC; Nieuwenhuis, Gilzenrat, Holmes, & Cohen, 2005), a minute structure (German et al., 1988) that projects to almost every region of the brain, with a special emphasis on areas involving attentional processing (Aston-Jones et al., 2000). The LC seems to be activated in response to the detection of a salient item. In addition, the effect of LC innervation includes amplification of excitatory responses for feature selective cells, and in a visual discrimination paradigm, monkey LC neurons were activated with a temporal profile that seems to match the AB. This observation prompted Nieuwenhuis et al.'s (2005) model; see Figure 25. The model can be divided into two components: the behavior network (comprising the input, decision, and detection layers) and the LC.

The behavior network is a simple feed-forward network, with one-to-one interlayer connections. Cross-talk connections are also included between input and decision layer nodes, reflecting feature similarity between stimuli. In addition, excitatory self-loops are included to sustain activation at decision and detection nodes. However, these loops are not strong enough to yield an active memory. Finally, as a reflection of its role, nodes in the decision layer compete through lateral inhibition.

The LC circuit modulates activity in the behavior network. Specifically, the LC is excited by detection of a salient stimulus, because target nodes in the decision layer project to the LC node. LC activity has a modulatory effect on the behavior network, simulating the release of norepinephrine. This release multiplicatively scales the afferent signals to network units, transiently adjusting their gain. It is important, though, that after firing, the LC enters a refractory period. It is unavailability of the LC during this period that causes the blink. At Lag 1, T2 benefits from the LC firing initiated by the T1, generating sparing.

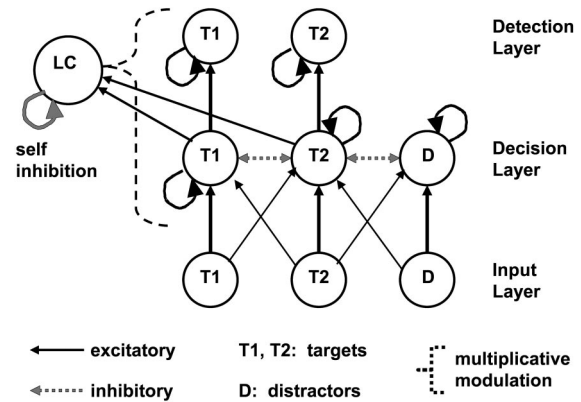


Figure 25. Depiction of the locus coeruleus (LC) model. Note that inhibitory and cross-talk connections between T1 and D are not shown for simplicity of presentation and point size of arrows indicates weight strength. This diagram is from "The Role of the Locus Coeruleus in Mediating the Attentional Blink: A Neurocomputational Theory," by S. Nieuwenhuis, M. S. Gilzenrat, B. D. Holmes, and J. D. Cohen, 2005, *Journal of Experimental Psychology: General*, 134, p. 295, Figure 2. Copyright 2005 by the American Psychological Association. Adapted with permission of the author.

Assessment and comparison to ST². Nieuwenhuis et al.'s (2005) LC model makes an important contribution to understanding the AB. A strength is that the model is framed within the context of a broad neurophysiological theory of attentional function. There are also important similarities between the approach and the ST² theory, most notably, between the LC and ST²'s TAE. Both are initiated by detection of a salient stimulus and their temporal characteristics are similar.

Nieuwenhuis et al.'s (2005) model is also consistent with a temporal, rather than sequential (by-item), interpretation of blink onset. That is, it suggests that the blink has a fixed temporal dynamic, the onset of which is not regulated by intervening distractors, as explored in Prediction 1 (see Section 5.1).⁹ This aspect of Nieuwenhuis et al.'s model suggests that it is also consistent with increased processing of the T1 + 1 slot (Chua et al., 2001). There are though differences between the approaches.

1. Perhaps most important, Nieuwenhuis et al.'s (2005) model suffers the problem we discussed in Section 6.3. In fact, in Nieuwenhuis et al.'s model, the length of the LC refractory period is not fixed. Rather, stronger LC firings lengthen the LC refractory period, which, in turn, means that greater bottom-up trace strength leads to a longer refractory period. This facet of Nieuwenhuis et al.'s model has the consequence that unmasking T1s (i.e., T1 + 1 blank conditions) would lengthen the blink. This

⁹ Indeed, the experimental findings in Nieuwenhuis et al. (2005) give added weight to our confirmation of Prediction 1.

Table 1
Commonalities Among Attentional Blink Models

Commonality	ST ²	Gated auto-associator	Global workspace	Conflict monitoring	CODAM	Locus coeruleus
Lateral inhibition between type representations	✗	✗	✓	✓ ^a	✓/✗ ^b	✓
TAE with refractory period	✓	✗	✗	✓	✗	✓
TAE is location specific	✓			✓		✗
TAE withheld during tokenization	✓			✗		✗
Target reps maintained to retrieval	✓	✓	✗	✗	✓/✗ ^c	✗
Active maintenance of entire type	✗	✗ ^d	✗	✗	✓/✗ ^c	✗
Maintenance of pointer to type (tokens)	✓	✗	✗	✗	✗	✗
Uses discrete episodic contexts (tokens)	✓	✗	✗	✗	✗	✗
Extraction of temporal order	✓	✗	✗	✗	✗	✗

Note. A ✓ denotes that the model has this feature, and a ✗ denotes that it does not.

ST² = simultaneous type, serial token; CODAM = corollary discharge of attention movement; TAE = transient attentional enhancement.

^a Lateral inhibition enforced only when conflict is high. ^b Depends on what are interpreted as types (object map or working memory nodes). ^c There is no reason why working memory nodes could not sustain until retrieval, although they do not seem to do so in the current implementation. ^d Weights are maintained, but activation is not.

stands contrary to the data, which, as previously discussed, show that unmasking attenuates the blink.¹⁰

2. ST²'s TAE is an additive enhancement, whereas the LC enhances by increasing the signal-to-noise ratio. The relative benefit of these two approaches is not yet clear. However, it is important to note that there is a gating aspect to Nieuwenhuis et al.'s (2005) mechanism, which is not present with ST²'s TAE.
3. In addition, although the TAE is assumed to be location specific, this is not the interpretation associated with the LC. This is an important point of comparison that should yield experiments to differentiate the two methods.
4. As acknowledged by Nieuwenhuis et al. (2005), their model suffers the common problem that (unlike ST²) nothing is actually sustained to retrieval. That is, T1 and T2 activations rise and fall at the detection layer with a time course in the range of a few hundred milliseconds of simulated time.
5. Finally, Nieuwenhuis et al.'s (2005) model does not reproduce as extensive a set of AB data as ST², and there is no handling of temporal order.

6.9. Synthesis Across Models

We now concisely summarize the AB modeling landscape. For this purpose, Table 1 shows commonalities across models. There is a further AB model (Barnard & Bowman, 2004), which we do not discuss because it is not a neural network.

7. ST² and Informal Theories of the AB

We now relate ST² to informal theories of the blink.

7.1. The Interference Theory

The interference theory (Shapiro, Arnell, & Raymond, 1997) is a prominent explanation of the AB. We discuss how our model relates to this theory now.

Serial Order

Current formulations of the interference theory do not consider serial order. It is assumed that the four relevant items (T1, T2, T1 + 1, and T2 + 1) are all placed in a buffer, and weights are associated with these items. However, these weights bias the competition between items in the buffer, rather than enabling serial order to be retrieved. It is though clear that some order information can be extracted from RSVP streams and thus that such information must be recorded somewhere and must somehow impact on retrieval.¹¹

However, in respect of ordering, perhaps the most critical difference between the ST² model and interference theory is that the latter seems to imply a sequential rather than temporal initiation of the blink: that is, that the arrival of the T1 + 1 item is the key initiator of the blink, as, for example, encapsulated in Battye's (2003) conflict monitoring model. Thus, if the T1 + 1 distractor is delayed, it is suggested that the blink profile will be slowed. This is in contrast to the ST² model, which, as previously emphasized, assumes that the blink profile is time rather than item dependent. Verification of Prediction 1 in Section 5.1 (and also the experi-

¹⁰ Although, it should be emphasized that this is a weakness that Nieuwenhuis et al. (2005) acknowledged and informally argued could be resolved in a revision of their current implementation.

¹¹ In fact, it should be noted that none of the prominent informal theories of the blink explicitly consider serial order. So, the interference theory should not be overly criticized in this respect.

mental findings in Nieuwenhuis et al., 2005) is consistent with the ST² approach in this respect.

Categorical Interference Versus Low-Level Masking

The manner in which the T1 + 1 and T2 + 1 items disrupt processing of their respective targets is a hotly debated topic. There are two extreme positions:

1. *Low-level masking.* The disruption is purely at the low level of visual features, that is, distractors following targets visually mask targets (Seiffert & Di Lollo, 1997).
2. *Categorical interference.* The disruption occurs at a categorical level, perhaps even at retrieval (Isaak, Shapiro, & Martin, 1999). Thus, the more categorically similar targets and target + 1 distractors are, the more confusable they are, and the more difficult it is to encode and/or retrieve the correct targets.

Our model selects 1 over 2. That is, it is assumed that weaker visual traces are the key cause of the AB. Masking weakens neural activation traces, thereby requiring more time to bind tokens at later stages. Recent data are generally consistent with the selection of 1 over 2 (Giesbrecht, Bischof, & Kingstone, 2004). Efforts to adjust blink depth by manipulating semantic similarity between T1 and the T1 + 1 distractor have generally not been successful.

Chun and Potter (1995) used mathematical symbols and found that categorically similar distractors caused deeper blinks than dissimilar distractors. However, it seems likely that the mathematical symbols used were not as effective visual masks as digits. In particular, Maki, Bussard, Lopez, and Digby (2003) attempted to replicate Chun and Potter's findings using false fonts that were matched for pixel density with digits and letters. From a categorical perspective, these false fonts should have been less similar to digits or letters than the symbols, yet no blink attenuation was observed. Their own experiments, using symbols as masks, did replicate the Chun and Potter findings by strongly attenuating the blink. The converse of this experiment (i.e., using letters as distractors and symbols as targets) failed to find an attenuated blink. These data stand against the categorical interference hypothesis, suggesting that visual featural masking properties of distractors are most important in determining blink depth.

7.2. Temporary Loss of Control (TLC)

An important new theory of the blink is based on the idea that T1 processing generates a period in which the system is vulnerable to a TLC (Di Lollo et al., 2005). Thus, a central processor is initially allocated to maintaining an input filter (similar to our salience filter) in a state in which it filters out distractors, allowing only targets to pass. However, T1 detection initiates a period in which the central processor is occupied with T1 stimulus processing and response planning. As a result, the central processor is unable to maintain the input filter, and the system becomes vulnerable to exogenous control. The arrival of a distractor item will then trigger a change in the system's configuration, causing it no longer to be optimally tuned to the target category. It is this reconfiguration that makes the system vulnerable to missing T2s.

The key findings that motivate the TLC theory are that sequences of three or even four consecutive RSVP targets seem immune to the blink (Di Lollo et al., 2005; Olivers, van der Stigchel, & Hulleman, 2005). In terms of the ST² model, this intriguing data suggest that TAE unavailability during tokenization is not absolute and that it may refire in response to sustained TFL activation. However, there are a number of empirical questions that need to be addressed before informed modeling of these data can be undertaken. First, Weichselgartner and Sperling (1987) in their RSVP study found two glimpses in a consecutive target paradigm; see Section 8.3. Why Weichselgartner and Sperling identified a blinklike effect when their participants had to encode four items in a row whereas Di Lollo et al. (2005) and Olivers et al. (2005) did not find such an impairment needs to be explored. The Weichselgartner and Sperling task is indeed somewhat different to Di Lollo et al.'s and Olivers et al.'s, but exactly why that changed the profile of data needs to be clarified.

In addition, in order to model Di Lollo et al.'s (2005) and Olivers et al.'s (2005) data, we need to know the profile of order errors among the three or four targets. This would indicate which items are bound to which tokens. Nobody has reported order information in three or four target experiments to date. Finally, there are existing data that current formulations of the TLC find difficult to explain. For example, the findings of our Prediction 1 (see Section 5.1) and the experiments contained in Nieuwenhuis et al. (2005) suggest that the blink is temporal, rather than sequential (i.e., by item). In contrast, the TLC theory would suggest that the first distractor after the T1 should initiate the blink; this is not what we find in our 50-ms SOA data.

7.3. Central Interference Theory (CIT)

ST² sits relatively well with Jolicoeur's (1998) CIT. This is not surprising because CIT has a good deal in common with Chun and Potter's (1995) two-stage model; for example, it assumes that a parallel first stage meets a sequential second stage, which is involved in encoding into short-term memory. However, Jolicoeur emphasized the spectrum of tasks that can interfere with such short-term memory encoding, for example, response selection, ongoing short-term memory encoding of a different item, and task switching. Furthermore, as a reflection of the claimed central nature of the bottleneck, it is asserted that the process of encoding into short-term memory is essentially amodal (Arnell & Jolicoeur, 1999).

A particularly strong commonality with ST² is that the CIT emphasizes how the ease of the T1 task affects the length of the blink, because the longer encoding of T1 into short-term memory takes, the more decayed the waiting T2 will be by the time it can be encoded. ST² provides a concrete realization of this position.¹² With regard to Arnell and Jolicoeur's (1999) claim that the bottleneck is central and essentially amodal, ST² is largely neutral. ST² is currently formulated in the visual modality, as a reflection of the data that we have been attempting to reproduce. However, a reformulation to accommodate an amodal locus of interference could be explored.

¹² In this respect, it is also interesting to note that the sort of crossover effects that Jolicoeur (1998) reported when comparing speeded and unspeeded T1 tasks have also been identified with ST² when the ease of the T1 task is varied (Wyble & Bowman, 2005).

DISCUSSION AND CONCLUDING REMARKS

8. ST² and Temporal Attention in General

8.1. Basic Principles

The ST² theory is a general framework for considering visual temporal attention and WM encoding. It is based on five principles.

1. *Two stages.* A cascaded parallel first stage is followed by a serial second stage that is closely tied to WM encoding.
2. *Salience filter.* A filter emphasizes salient and deemphasizes nonsalient items.
3. *Types–tokens.* The association of a token with an active type is the key mechanism involved in ST² WM encoding.
4. *TAE.* In response to the detection of a salient stimulus, the TAE provides a brief (but spatially specific) enhancement of type representations.
5. *Binding pool.* Through the binding pool, types and tokens can be arbitrarily associated, order information can be recorded, and degenerate type–token bindings can be represented; for example, multiple types can be bound into the same token. In addition, because type resources are freed up once an item has been tokenized, further instances of the same item can be perceived and encoded.

As evidence of the success of the ST² framework, we show how it satisfies the five functional requirements we set out for our theory in Section 1.2.

1. *Salience detection.* ST²'s salience filter enables rapid detection of salient stimuli that are briefly presented in demanding visual environments.
2. *Sustain representations.* The TAE sustains fleeting, but salient, representations, enabling encoding into WM.
3. *Ascribe episodic context.* The types–tokens system enables discrete episodic contexts to be ascribed to items.
4. *WM maintenance.* Because multiple tokens are available, many items can be maintained together in WM.
5. *Repetitions.* Multiple tokens can be bound to the same type, compactly maintaining multiple instances of the same item in WM.

We have also provided neural-ST², a connectionist realization that is faithful to the principles of the ST² theory. In respect of neural network modeling, our approach is relatively high level. In particular, we have not modeled individual biological neurons, but rather, nodes in our model correspond to assemblies of (perhaps many thousands of) biological neurons. This is an appropriate level of abstraction to work at for a number of reasons. First, the main constraints on our modeling are cognitive-level behavioral data; in particular, single-cell recordings in such a setting are not available. Second, our theoretical position is most naturally elaborated in terms of high-level neural dynamics; it is unclear that low-level details of biological implemen-

tation would be illuminating. In particular, the systems level at which we have worked has enabled us to develop a relatively simple and computationally tractable instantiation of our theory. Finally, the level of modeling we work at corresponds to that used in a large number of connectionist models and has been convincingly justified (O'Reilly & Munakata, 2000).

In terms of theories of attention, it is also interesting to note that we have not required any direct analogue of biased competition (Desimone & Duncan, 1995) in ST². This could be because in the temporal attention setting, items compete in time, rather than space. Biased competition usually has a spatial focus.

8.2. Tokens, Object Continuity, and Object Files

The ST² framework takes a good deal of inspiration from Chun's (1997) article. By exploring AB and repetition blindness effects within a single RSVP setup, Chun articulated an informal types–token theory of the AB. This theory made the important link between T1 tokenization and the blink: the link that during the blink, the T2 type was locked out of WM encoding while the T1 type was being associated with a token. However, there are a number of respects in which ST² extends and diverges from Chun's theory. First, ST² makes the strong claim that the central role of distractors in generating the blink is to weaken target traces through masking. In contrast, Chun argued for higher level categorical similarity effects. See Section 7.1 for a justification of the position we have taken on this topic. Second, Chun made a distinction between object tokens and spatiotemporal tokens; this distinction has not been required in the current formulation of ST². Third, and most important, the link between TAE availability and tokenization that in ST² is actually the direct mechanism for generating the blink was not present in Chun (1997). In this sense, ST² elaborates an exact mechanism by which the gate can be closed on T2 while T1 is being tokenized.

A number of recent studies have revealed a tantalizing effect of object constancy on the AB (Kellie & Shapiro, 2004; Raymond, 2003). The central message here is that if there is object constancy across the items of an RSVP stream—for example, the items represent the gradual morphing of one object into another (Kellie & Shapiro, 2004)—then targets marked by small featural changes within the stream will not generate a blink. These findings suggest a link between the AB and object file updating. That is, in a stream with object constancy, a single object file is set up, which is continually updated as the object changes. However, it is important to note that these object updates, even when target discriminating features are updated, do not generate a blink. Rather, it is the creation of new object representations that generates a blink.¹³

Although ST²'s tokens are not currently as representationally rich as object files, the two concepts clearly have similarities. For example, modulo the Lag 1 case, when presented with an AB stream, ST²'s tokens will come to represent distinct objects, with the featural attributes of those objects associated with tokens through the binding pool. Thus, it could be that the token binding

¹³ These findings resonate with studies showing that discriminating the duration of an event in a (singly perceived) RSVP stream does not generate a blink (Sheppard, Duncan, Shapiro, & Hillstrom, 2002). Presumably, gap judgments do not initiate the generation of new object representations.

and handover mechanism implemented in ST² should be viewed as the creation of a new WM object. This is the subject of ongoing research.

8.3. Sperling and Coworkers' Theory of Temporal Attention

A prominent and powerful theory of temporal attention is that developed by Sperling and coworkers (Shih & Sperling, 2002; Sperling & Weichselgartner, 1995; Weichselgartner & Sperling, 1987). They investigated tasks in which a window of attentional engagement is cued in which a sequence of (usually four) items presented in RSVP had to be reported. The bimodal profile of temporal performance obtained by Weichselgartner and Sperling's (1987) Procedure 2 is particularly noteworthy. They identified a first glimpse in which the first and second items after the cue were well reported; this was followed by significantly impaired performance for the third item and then a second glimpse in which performance on the fourth and fifth items recovered. The ST² theory would explain this profile in terms of two token-blast episodes separated by a window of blaster unavailability.¹⁴

Sperling and coworkers' (Shih & Sperling, 2002; Sperling & Weichselgartner, 1995; Weichselgartner & Sperling, 1987) theory seeks to explain trajectories between location-specific attentional episodes. Although it does not explicitly address spatial relocation of attention, ST² explains temporal attention in terms of discrete token allocation events, which resonates with Sperling and coworkers' emphasis on discrete attentional episodes. In addition, ST²'s TAE plays a similar role to Sperling and coworkers' attentional gate, which opens in response to detection of a cue. However, Sperling and coworkers' attentional gate is multiplicative, whereas ST²'s TAE is additive. Most significant though, the ST² theory assumes a close coordination between tokenization and TAE availability, whereby the TAE fires once and then is held offline during tokenization. This TAE-tokenization interaction is central to ST²'s explanation of the blink and is not present in Sperling and coworkers' theory.

Finally, Sperling and coworkers' (Shih & Sperling, 2002; Sperling & Weichselgartner, 1995; Weichselgartner & Sperling, 1987) models propose that temporal order can be extracted from the activation strengths with which items are encoded into WM. This approach would suggest that at Lag 1, T1–T2 swaps would be higher than chance, because T2 is typically reported better than T1 at this lag (Potter et al., 2002). However, swaps are below chance at Lag 1; see Figure 15d. Such data have motivated our token-based approach to temporal order encoding.

8.4. Concluding Remark

As is in fact the case with all models, there is no sense to which the ST² model is in an absolute sense the right model of temporal attention. However, we do believe that the ST² model illustrates a set of issues that would need to be addressed by any competitor model. Most prominent among these is the types–tokens distinction. Some mechanism is needed by which an instance-specific marker of the occurrence of a type is held in WM. Furthermore, such a marker needs to record order information, accommodate repetition of types (beyond the repetition blindness window), and

enable types to be regenerated during retrieval. Our binding pool satisfies these criteria in a neurally plausible manner.

¹⁴ Although, note that Di Lollo et al. (2005) and Olivers et al. (2005) failed to obtain separate glimpses when they considered sequences of three or more targets in an RSVP task; see Section 7.2.

References

- Amit, D. J., Bernacchia, A., & Yakovlev, V. (2003). Multiple-object working memory—A model for behavioral performance. *Cerebral Cortex*, 13, 435–443.
- Anderson, A. K., & Phelps, E. A. (2001, May 17). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature*, 411, 305–309.
- Arnell, K. M., & Jolicoeur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 630–648.
- Aston-Jones, G., Rajkowski, J., & Cohen, J. (2000). Locus coeruleus and regulation of behavioral flexibility and attention. *Progress in Brain Research*, 126, 165–182.
- Bal, T., & McCormick, D. A. (1993). Mechanisms of oscillatory activity in guinea-pig nucleus reticularis thalami in vitro: A mammalian pacemaker. *Journal of Physiology*, 468, 669–691.
- Barnard, P. J., & Bowman, H. (2004). Rendering information processing models of cognition and affect computationally explicit: Distributed executive control and the deployment of attention. *Cognitive Science Quarterly*, 3, 297–328.
- Battye, G. (2003). *Connectionist modelling of attention and anxiety*. Unpublished doctoral dissertation, Medical Research Council's Cognition and Brain Sciences Unit, Cambridge, England.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.
- Chartier, S., Cousineau, D., & Charbonneau, D. (2004). A connectionist model of the attentional blink effect during a rapid serial visual task. In *Proceedings of the Sixth International Conference on Cognitive Modeling* (pp. 64–69). Mahwah, NJ: Erlbaum.
- Chua, F. K., Goh, J., & Hon, N. (2001). Nature of codes extracted during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 1229–1242.
- Chun, M. M. (1997). Types and tokens in visual processing: A double dissociation between the attentional blink and repetition blindness. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 738–755.
- Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 109–127.
- Chun, M. M., & Potter, M. C. (2000). The attentional blink and task switching within and across modalities. In K. Shapiro (Ed.), *Temporal constraints in human information processing* (pp. 20–35). Oxford, England: Oxford University Press.
- Cohen, J. D., & Huston, T. A. (1994). Progress in the use of interactive models for understanding attention and performance. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and non-conscious information processing* (pp. 453–476). Cambridge, MA: MIT Press.
- Coltheart, M. (1983). Iconic memory. *Philosophical Transactions of the Royal Society London, Series B*, 302, 283–294.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and

- stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3, 201–215.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–114.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, 112, 3–42.
- Davelaar, E. J., & Usher, M. (2004). An extended buffer model for active maintenance and selective updating. In H. Bowman & C. Labiouse (Eds.), *Proceedings of the 8th Neural Computation and Psychology Workshop* (Vol. 15, pp. 3–14). Singapore: World Scientific.
- Deco, G., Rolls, E. T., & Horowitz, B. (2004). “What” and “where” in visual working memory: A computational neurodynamical perspective for integrating fMRI and single-neuron data. *Journal of Cognitive Neuroscience*, 16, 683–701.
- Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences, USA*, 100, 8520–8525.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, USA*, 93, 13494–13499.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.
- Di Lollo, V., Kawahara, J., Shahab Ghorashi, S. M., & Enns, J. T. (2005). The attentional blink: Resource depletion or temporary loss of control? *Psychological Research*, 69, 191–200.
- Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, 92, 53–78.
- Duncan, J. (1981). Directing attention in the visual field. *Perception & Psychophysics*, 30, 90–93.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise-letters on identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16, 143–149.
- Fell, J., Dietl, T., Grunwald, T., Kurthen, M., Klaver, P., Trautner, P., et al. (2004). Neural bases of cognitive ERPs: More than phase reset. *Journal of Cognitive Neuroscience*, 16, 1595–1604.
- Fragopanagos, N., Kockelkoren, S., & Taylor, J. G. (2005). A neurodynamic model of the attentional blink. *Cognitive Brain Research*, 24, 568–586.
- German, D. C., Walker, B. S., Manaye, K., Smith, W. K., Woodward, D. J., & North, A. J. (1988). The human locus coeruleus: Computer reconstruction of cellular distribution. *Journal of Neuroscience*, 8, 1776–1788.
- Giesbrecht, B., Bischof, W. F., & Kingstone, A. (2003). Visual masking during the attentional blink: Tests of the object substitution hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 238–258.
- Giesbrecht, B., Bischof, W. F., & Kingstone, A. (2004). Seeing the light: Adapting luminance reveals low-level visual processes in the attentional blink. *Brain and Cognition*, 55, 307–309.
- Giesbrecht, B., & Di Lollo, V. (1998). Beyond the attentional blink: Visual masking by object substitution. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1454–1466.
- Hebb, D. O. (1949). *The organization of behaviour*. New York: Wiley.
- Hotson, J. R., & Prince, D. A. (1980). A calcium-activated hyperpolarization follows repetitive firing in hippocampal neurons. *Journal of Neurophysiology*, 43, 409–419.
- Hummel, J. E., & Biederman, L. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480–517.
- Isaak, M. I., Shapiro, K. L., & Martin, J. (1999). The attentional blink reflects retrieval competition among multiple rapid serial visual presentation items: Tests of an interference model. *Journal of Experimental Psychology: Human Perception & Performance*, 25, 1774–1792.
- Jolicoeur, P. (1998). Modulation of the attentional blink by on-line response selection: Evidence from speeded and unspeeded task decisions. *Memory & Cognition*, 26, 1014–1032.
- Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition*, 27, 117–143.
- Kanwisher, N. G. (1991). Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 404–421.
- Kellie, F. J., & Shapiro, K. L. (2004). Object file continuity predicts attentional blink magnitude. *Perception & Psychophysics*, 66, 692–712.
- Keyser, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences*, 6, 120–125.
- Koch, C. (1999). *Biophysics of computation: Information processing in single neurons*. Oxford University Press.
- Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward masking. *Proceedings of the National Academy of Sciences, USA*, 92, 5587–5591.
- Kranczioch, C., Debener, S., Schwarzbach, J., Goebel, R., & Engel, A. K. (2005). Neural correlates of conscious perception in the attentional blink. *NeuroImage*, 24, 704–714.
- Kristjansson, A., Mackeben, M., & Nakayama, K. (2001). Rapid, object-based learning in the deployment of transient attention. *Perception*, 30, 1375–1387.
- Kristjansson, A., & Nakayama, K. (2003). A primitive memory system for the deployment of transient attention. *Perception & Psychophysics*, 65, 711–724.
- Liaw, J. S., & Berger, T. W. (1996). Dynamic synapse: A new concept of neural representation and computation. *Hippocampus*, 6, 591–600.
- Logan, G. D. (2004). Cumulative progress in formal theories of attention. *Annual Review of Psychology*, 55, 207–234.
- Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996, October 17). Word meanings can be accessed but not reported during the attentional blink. *Nature*, 383, 616–618.
- Maki, W. S., Bussard, G., Lopez, K., & Digby, B. (2003). Sources of interference in the attentional blink: Target-distractor similarity revisited. *Perception & Psychophysics*, 65, 188–201.
- Maki, W. S., Couture, T., Frigen, K., & Lien, D. (1997). Sources of the attentional blink during rapid serial visual presentation: Perceptual interference and retrieval competition. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1393–1411.
- Marois, R., Do-Joon, Y., & Chun, M. M. (2004). The neural fate of consciously perceived and missed events in the attentional blink. *Neuron*, 41, 465–472.
- McClelland, J. L. (1993). Toward a theory of information-processing in graded, random and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Cambridge, MA: MIT Press.
- Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, 16, 5154–5167.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7, 134–140.
- Muller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 315–330.
- Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, 29, 1631–1647.
- Nieuwenhuis, S., Gilzenrat, M. S., Holmes, B. D., & Cohen, J. D. (2005). The role of the locus coeruleus in mediating the attentional blink: A

- neurocomputational theory. *Journal of Experimental Psychology: General*, 134, 291–307.
- Nieuwenstein, M. R., Chun, M. M., van der Lubbe, R. H. J., & Hooge, I. T. C. (2005). Delayed attentional engagement in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1463–1475.
- Olivers, C. N., van der Stigchel, S., & Hulleman, J. (2005). Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research*, 8, 1–14.
- O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.
- Pantic, L., Torres, J. J., Kappen, H. J., & Gielen, S. C. (2002). Associative memory with dynamic synapses. *Neural Computation*, 14, 2903–2923.
- Passingham, D., & Saka, K. (2004). The prefrontal cortex and working memory: Physiology and brain imaging. *Current Opinion in Neurobiology*, 14, 163–168.
- Petrides, M. (1995). Impairments on nonspatial self-ordered and externally ordered working memory tasks after lesions of the mid-dorsal part of the lateral frontal cortex in the monkey. *Journal of Neuroscience*, 15, 359–375.
- Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, 21, 156–161.
- Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 979–992.
- Potter, M. C., Staub, A., & O'Connor, D. H. (2002). The time course for competition for attention: Attention is initially labile. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1149–1162.
- Raymond, J. E. (2003). New objects, not new features, trigger the attentional blink. *Psychological Science*, 14, 54–59.
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, 18, 849–860.
- Rolke, B., Heil, M., Streb, J., & Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology*, 38, 165–174.
- Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, 11, 300–311.
- Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, 8, 363–370.
- Seiffert, A. E., & Di Lollo, V. (1997). Low-level masking in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 1061–1073.
- Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science*, 16, 114–122.
- Shapiro, K. L., Arnell, K. M., & Raymond, J. E. (1997). The attentional blink. *Trends in Cognitive Sciences*, 1, 291–297.
- Shapiro, K. L., Caldwell, J., & Sorensen, R. E. (1997). Personal names and the attentional blink: A visual “cocktail party” effect. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 504–514.
- Shapiro, K. L., Driver, J., Ward, R., & Sorensen, R. E. (1997). Priming from the attentional blink: A failure to extract visual tokens but not visual types. *Psychological Science*, 8, 95–102.
- Sheppard, D. M., Duncan, J., Shapiro, K. L., & Hillstrom, A. P. (2002). Objects and events in the attentional blink. *Psychological Science*, 13, 410–415.
- Shih, S. I., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review*, 109, 260–305.
- Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review*, 102, 503–532.
- Taylor, J. G. (2002). Paying attention to consciousness. *Trends in Cognitive Sciences*, 6, 206–210.
- Taylor, J. G., & Rogers, M. (2002). A control model of the movement of attention. *Neural Networks*, 15, 309–326.
- Treisman, A. M. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society London, Series B*, 353, 1295–1306.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1656–1674.
- von der Malsburg, C. (1981). *The correlation theory of brain function* (Technical Report No. 81–2). Gottingen, Germany: Max-Planck Institute for Biophysical Chemistry.
- Weichselgartner, E., & Sperling, G. (1987, November 6). Dynamics of automatic and controlled visual attention. *Science*, 238, 778–780.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4, 1120–1135.
- Wyble, B., & Bowman, H. (2005). Computational and experimental evaluation of the attentional blink: Testing the simultaneous type serial token model. In B. G. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *CogSci 2005: XXVII Annual Conference of the Cognitive Science Society* (pp. 2371–2376). Mahwah, NJ: Erlbaum.
- Wyble, B., & Bowman, H. (2006). A neural network account of binding discrete items into working memory using a distributed pool of flexible resources [Abstract, electronic version]. *Journal of Vision*, 6, 33a.

(Appendix follows)

Appendix

Implementation and Experimental Methods

A.1. Parameter Settings

The following parameters are fixed across all neural-ST² layers: (from Equation 1) $DT_{VM} = 1.2$, $EE = 3$, $EI = -0.2$, $EL = 0$, and (from Equation 2) $\gamma = 5$. Table A1 gives parameter settings per layer. Different connectivity types arise in neural-ST²; see Panel A of Figure A1. Table A2 uses these types to detail interlayer connectivity, in which layer numbering is as per Table A1. We illustrate connections between elements of the token binding process directly, with the value of weights explicitly shown; see Panel B Figure A1.

A.2. Blaster Implementation

The blaster implementation is shown in Figure A2. All connections in this module saturate at a very low level, so the value of nodes, once a threshold is crossed, is essentially constant with increasing activation. The blaster implementation comprises two

Table A1
Parameter Settings Per Layer

Layer	Leak	No. of nodes
1. Input	.07	12
2. Masking	.01	12
3. Item	.04	12
4. Item off	.02	12
5. TFL	.02	12
6. TFL off	.07	12
7. Binder gate	.07	6
8. Binder trace	.01	6
9. WM gate	.07	3
10. WM trace	.01	3
11. Blaster input on	.07	1
12. Blaster input off	.07	1
13. Blaster output on	.07	1
14. Blaster output off	.07	1

Note. TFL = task filtered layer; WM = working memory.

Table A2
Connectivity Between Layers

Connection	Value	Type	Threshold
Excitatory			
1 → 2	.022	1	0.15
2 → 3	.014	1	0.15
3 → 5 ^a	.015	1	0.15
3 → 3	.0095	1	0.15
4 → 4	.0095	1	0.4
6 → 6	.01	1	0.4
5 → 5	.022	1	0.4
5 → 6	.0075	1	0.52
3 → 4	.02	1	0.52
9 → 10	.0055	1	0.4
10 → 10	.06	1	0.15
5 → 11	.018	2	0.15
13 → 5 ^b	.8	2	0.48
13 → 3 ^b	.3	2	0.48
11 → 13 ^b	4	1	0.15
13 → 14	.5	1	0.7
7 → 12	.01	2	0.4
7 → 8	.0039	1	0.52
8 → 8	.01	1	0.15
11 → 12 ^b	5	1	0.15
13 → 13	.04	1	0.15
12 → 12	.01	1	0.15
Inhibitory			
1 → 2	-.105	3	0.15
2 → 2	-.06	3	0.15
6 → 5	-.12	1	0.4
4 → 3	-.15	1	0.52
10 → 9	-.6	1	0.4
12 → 11	-1	1	0.15
14 → 13	-1	1	0.15

Note. Value is the weight setting, type the connectivity type (see Figure A1), and threshold the value of θ used in the presynaptic output equation. The item off and task filtered layer off nodes are weakly self-excitatory ($4 \rightarrow 4$, $6 \rightarrow 6$, and $12 \rightarrow 12$), which was not discussed in the *Type Processing: The Item Layer and The TFL* subsection of Section 4.3 or, in Section A.2.

^a Saturating connection: Output stops at .2, limiting output dynamic range. ^b Saturating connection: Output stops at .01, limiting output dynamic range.

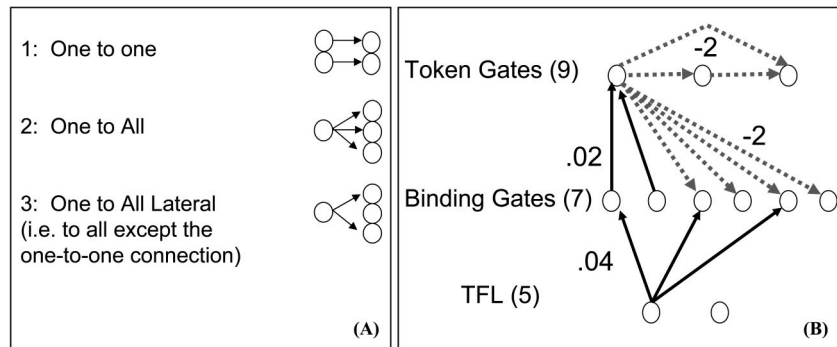


Figure A1. A: Connection types. B: Connectivity into and out of the binding gates; connectivity shown for one neuron is repeated across that layer. Solid links are excitatory, and dashed links are inhibitory. TFL = task filtered layer.

on-off circuits, distinguished as blaster input, which governs initiation of the blaster, and blaster output, which governs the blaster's output effect. Thus, the blaster is initiated by TFL excitation of the blaster input on node (via the projection labeled a). However, the resulting activation is short-lived and rapidly curtailed when the blaster input off node feeds inhibition back onto the blaster input on node (generating the blaster refractory period). Furthermore, this suppression is maintained by ongoing binding, via the projection labeled c.^{A1}

The resulting brief on node activation of the blaster input circuit has the role of instigating blaster output activation through the link labeled d. This output circuit generates a temporally fixed output to the TFL and item layer (via the projection labeled b). This effect is obtained through the blaster output on node having different thresholds for outputting along different connections, where Threshold 1 is less than 2 is less than 3.

When the blaster output on node is excited, it begins its self-excitatory cycle with a low Threshold 1, causing it to ramp up with a predictable time course. When activation crosses Threshold 2, it begins outputting at a fixed level over projection b. On crossing Threshold 3, it strongly excites the blaster output off node, which in turn inhibits it, ending the blaster's output at the appropriate time.

A.3. Experimental Methods

SOA Experiments

Fifty-four milliseconds SOA Experiments. Prediction 1 (see Section 5.1) involves perception of letter targets in a stream of digit distractors. MATLAB 6.5 and the psychophysics toolbox^{A2} were used to present RSVP streams to participants on a computer running Windows 2000. Display timing rates and screen refresh synchrony were confirmed with photodiodes. Participants were positioned in front of a 17-in. (43.18-cm) computer screen, displaying a white background, upon which black letters in 180-point Arial font were used as targets. Black digits in 220-point Arial font were used as distractors. The letters *I, M, O, Q, S, W, X,* and *Z* as well as digits *1* and *0* were excluded from all sets. Digits were 5-cm tall and approximately 3-cm wide, resulting in a viewing angle of $7.1^\circ \times 4.3^\circ$. Letters were 82% of this size.

Participants were University of Kent students, with normal or corrected-to-normal vision, who spoke English fluently. Ten participants undertook this experiment. Items were presented at 54-ms SOA, followed immediately by the next item. RSVP streams were 40 items long. Each trial was prefaced by a 500-ms fixation cross in the location of the stream. T1 could appear in Slots 9–17, and T2 could appear at Even Lags 2–16 afterward. At least eight distractors followed T2. Participants saw 96 two-target trials and 10 catch trials (with only one target) per block for three blocks for a total of 288 critical trials. Instructions presented on each trial asked participants to “Enter all of the letters you saw in order, then press Enter.” Participants were allowed to correct their input with backspace and were not given feedback. They were encouraged to guess if they were not sure but not to guess blindly. Trials were considered correct if participants got the correct identity, without regard to temporal order.

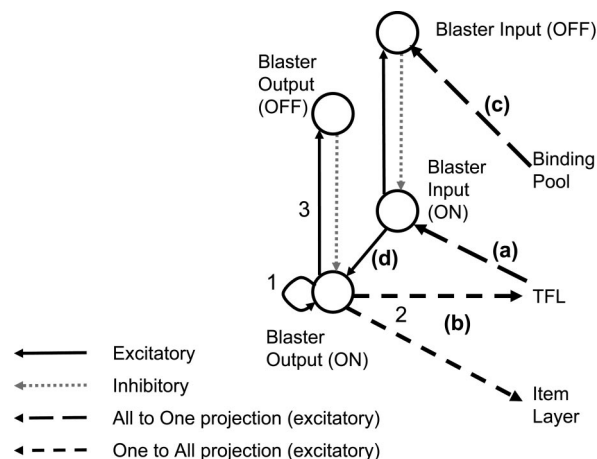


Figure A2. Full blaster implementation. Output connection thresholds emanating from the blaster output (on) node are denoted 1, 2, and 3, with 1 less than 2 less than 3. Note that connections a, b, and c are as per Figure 9. TFL = task filtered layer.

Ninety-four milliseconds SOA Experiments. The comparison data in Figure 19 were a subset of the results collected in a second experiment that was similar to the above. Fourteen University of Kent students participated. Items were presented at 94-ms SOA, followed immediately by the next item. RSVP streams were 18 items in length. T1 could appear in Slots 5–8, and T2 could appear from Lags 1–8 afterward. At least four distractors followed T2.

With equal probability, items could have a blank in the T1 + 1 slot or the T2 + 1 slot or no blank at all. Trial types were crossed in an 8×3 paradigm, with lag and blanks as primary factors. There were 144 trials per block with three blocks per participant and 14 catch trials with no T2. The T1 + 1 blank and T2 + 1 blank trials were discarded for this comparison.

Letter Pairs Experiment

On each trial, an RSVP stream was presented at 115 ms/item. Each stream consisted primarily of a sequence of digit trios (147, 258, 369, 470, 581, 692, 814, 703, and 925), presented in 100-point Arial font. Embedded within this stream were one or two targets selected from a prepared set of almost 12,000 possible quadruples of letters.^{A3} Characters were 2.4 cm in height. Digit trios were approximately 5.5 cm in width and

^{A1} Note this projection is depicted as inhibitory in Figure 9. But, in fact, it is realized as an excitatory projection onto the blaster input off node, which overall yields an inhibitory effect of the binding pool onto the blaster input on node.

^{A2} See <http://psychtoolbox.org>.

^{A3} These letters were chosen by scanning a corpus of over 150 assorted English novels including contemporary as well as classical authors and identifying the set of four-letter combinations that were the least likely to appear asymmetrically, either in order or in pairing. Although asymmetries may have existed in other languages or via popular acronyms, targets were always randomly chosen from this set so there could not have been a systematic bias.

letter pairs approximately 4.4 cm. Viewing angles were approximately 3.4° in height × 7° in width. Twelve University of Kent students participated in this experiment.

For the two targets in the stream, no repeats of individual letters were permitted. T1 could appear from Lags 4 to 9; T2, if present, would follow T1 by Lags 1–8; and there were at least three distractors following T2. Each participant viewed three blocks

consisting of 96 two-target trials and 10 catch trials for a total of 288 critical trials. After each trial, participants were asked first to “Enter the first pair of letters you saw in order, then press Enter” and the same for the second pair.

Received March 25, 2005

Revision received August 2, 2006

Accepted August 3, 2006 ■

Call for Papers *Journal of Experimental Psychology: Learning, Memory, and Cognition*
Special Section on Source Memory: Integrating Cognitive Behavioral and Cognitive Neuroscience Approaches

The *Journal of Experimental Psychology: Learning, Memory, and Cognition* invites manuscripts for a special section on source memory, to be compiled by guest editors Marcia K. Johnson and Mieke H. Verfaellie, working together with journal Associate Editor John Dunlosky. The goal of the special section is to showcase high-quality research that brings together behavioral, neuropsychological, and neuroimaging approaches to understanding the cognitive and neural bases of source memory. We are seeking cognitive behavioral studies that integrate cognitive neuroscience findings in justifying hypotheses or interpreting results and cognitive neuroscience studies that emphasize how the evidence informs cognitive theories of source memory. In addition to empirical papers, focused review articles that highlight the significance of cognitive neuroscience approaches to cognitive theory of source memory are also appropriate.

The submission deadline is June 1, 2007. The main text of each manuscript, exclusive of figures, tables, references, or appendixes, should not exceed 35 double-spaced pages (approximately 7,500 words). Initial inquiries regarding the special section may be sent to John Dunlosky (jdunlosk@kent.edu), Marcia K. Johnson (marcia.johnson@yale.edu), or Mieke H. Verfaellie (verf@bu.edu). Papers should be submitted through the regular submission portal for *JEP: Learning, Memory, and Cognition* (<http://www.apa.org/journals/xlm/submission.html>) with a cover letter indicating that the paper is to be considered for the special section. For instructions to authors and other detailed submission information, see the journal Web site at <http://www.apa.org/journals/xlm>.