Fortunate Conjunctions Revived: Feature Binding with the 2f-ST^2 Model

Srivas Chennu (srivas@gmail.com)
Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK

Howard Bowman (H.Bowman@kent.ac.uk)
Centre for Cognitive Neuroscience and Cognitive Systems, University of Kent, Canterbury, UK

Brad Wyble (bwyble@gmail.com)
Department of Psychology, Syracuse University, Syracuse NY, USA

Abstract

Temporal feature binding in vision refers to the process by which features of objects presented one after the other at the same spatial location are correctly bound together. In this paper, we describe a computational model of putative neural mechanisms that would produce this behaviour. These simulations highlight the role of transient attentional enhancement in mediating the temporal binding of features into working memory. This model builds upon previous approaches, and explains a range of behavioural findings relating to the patterns of illusory conjunctions observed in experiments. Further, it provides a parsimonious account of a counter-intuitive pattern of reaction time data.

Keywords: Binding Problem, Temporal Feature Binding, Neural Modelling

Introduction

We generally see the world as containing coherent bound objects. We might perceive the coincidence of the colour red and a rapidly moving form as a speeding red Ferrari. Such feature binding is effectively pre-conscious, and as such, we would not notice it as a task to be performed. However, the brain has to conjoin features in the environment in order that we can perceive bound objects. In other words, the brain has to solve the "binding problem". This refers to the cognitive and neural mechanisms by which feature binding is accomplished (Treisman, 1996). In experiments that impede the focusing of attention on specific visual objects, participants can be induced to produce errors in binding, generally referred to as *conjunction errors*. Under such circumstances, they often consciously perceive *illusory conjunctions* of visual features, which are defined as 'miscombination of features actually presented in a single display' (Treisman & Gelade, 1980). Since the pioneering work by Treisman and Gelade, the various aspects of feature binding in space have been extensively explored in numerous experiments.

Temporal Feature Binding Although largely unnoticed by us, the brain also has an impressive capacity to correctly bind features through time. For example, when we are driving along a busy road, it does a good job of tracking a large number of moving objects. However, such dynamic visual environments also cause us to make occasional errors. These errors in temporal binding occur when a feature from one object is mistakenly bound to a temporally neighbouring object. The pattern of these errors is revealing of the means by which the brain binds through time. Temporal binding errors can be generated in experiments using rapidly presented sequential stimuli, but have received relatively little theoretical attention. In response, this paper focuses on a theoretical exploration of temporal feature binding in vision. In particular, we build upon previous research to propose the *2f-ST*^{2} model of temporal feature binding (see Figure 1a). This neural network model emphasises the role of *transient attentional enhancement* as a mediator between two stages of visual information processing, enabling the binding of stimulus features into working memory (see Figure 1b).

Experiments involving feature binding in time usually employ the *Rapid Serial Visual Presentation* (RSVP) paradigm (see Figure 1b for an example), where a stream of stimuli is presented, with each rapidly replacing the previous one, at the rate of approximately 10/sec. The task-defining feature, present only in the target, is called the *key feature*. The feature of the target that is to be identified and reported is termed the *response feature*. For example, in Figure 1b, each RSVP trial consists of a stream of coloured letters, and participants are asked to detect the identity (the response feature) of the letter presented in red (the key feature). In such situations, multiple stimuli may be simultaneously processed in the visual system, leading to the possibility of the formation of illusory conjunctions. In this context, the visual system has the task of solving the binding problem in time: it must pick out the features of targets amongst multiple, temporally overlapping features of distractors. This process is referred to as temporal feature binding.

The BBS Model The BBS model (Botella, Barriopedro, & Suero, 2001) is currently the dominant account of temporal feature binding in vision, and indeed, it provides an accurate fit to available data. This model assumes that such binding occurs either through an *attentional focusing* (AF) route or through a sophisticated guessing (SG) route. These two binding routes are additional to the two parallel pathways posited for processing key and response features. The AF route is taken when attention has enough time to correctly pick out the target's key and response features. SG, on the other hand, comes into play when there is temporal uncertainty about the target's response feature. Importantly, AF leads to a distinct (always correct) outcome, requiring a binding route functionally different from that which produces conjunction errors. To justify this dual-route approach, the authors point to reaction time (RT) data from Botella (1992), which suggests that most

correct reports are processed earlier than conjunction errors, and hence are not 'fortunate conjunctions'. They suggest that most correct reports are produced by the fast and deterministic AF route. Conjunction errors, on the other hand, are produced by the slow and probabilistic SG route. However, the model leaves unexplained the mechanistic basis for the choice of binding route. Specifically, the BBS model incorporates a probability of taking the AF or the SG route, which is determined by the data being modelled; but it does not elaborate on how this probability might emerge from the nature of the stimuli being processed. Our model (2f-ST^2) avoids this limitation, since it describes how qualitatively different behavioural outcomes (correct reports vs. conjunction errors) might be realised by a single binding mechanism under the influence of variation in stimulus latency and strength.

The 2f-ST^2 Model

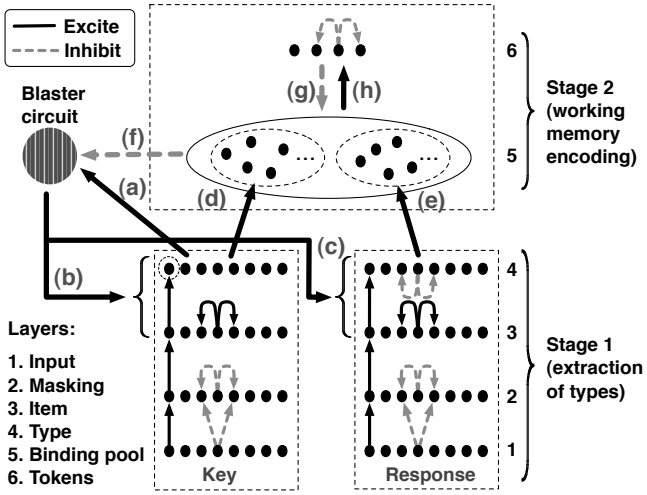
The 2-feature simultaneous type serial token (2f-ST^2) model has been designed to simulate temporal feature binding in RSVP experiments. It draws upon a previous model (Bowman & Wyble, 2007), and is, we argue, more functionally complete than the BBS model (Botella et al., 2001).

Broadly speaking, the architecture of the 2f-ST^2 model can be divided into two stages of processing (see Figure 1a). A parallel Stage 1 is responsible for extracting information (called *types* (Bowman & Wyble, 2007)) from RSVP items along multiple feature dimensions (e.g., item colour and shape). These dimensions are processed separately and concurrently, within two distinct, parallel pathways: the *key* and *response* pathways. Stage 2 consists of the *binding pool* and a set of *tokens* (Bowman & Wyble, 2007). The binding pool stores an association between a pair of type nodes (one from each of the two pathways) and a working memory token. Temporal coordination of these three nodes is orchestrated by a temporal attention mechanism, the *blaster*, which is triggered by activation in the key pathway.

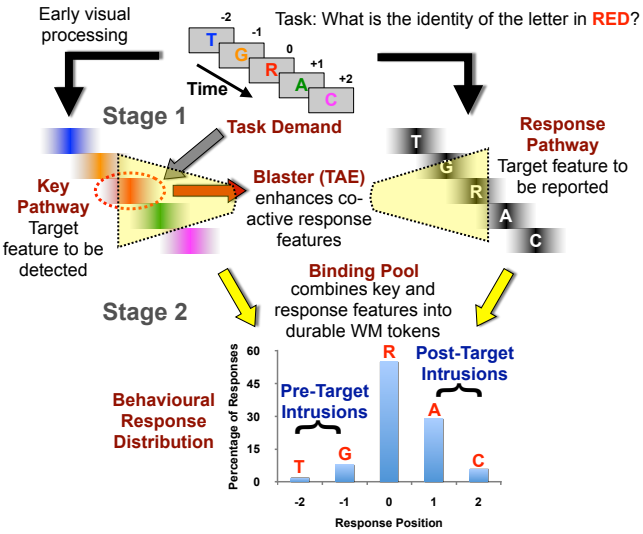
Key Pathway The constituent key features of items are extracted in the key pathway. At the key Type layer, task demand (depicted in Figure 1a as a dashed circle around the target type node) enhances type representations of target key features (the colour red in Figure 1b), and suppresses distractor key features. Hence, distractors are RSVP items that lack the target-defining key feature.

Blaster The blaster is triggered by activation from key target type nodes, via the connection labelled (a) in Figure 1a. Once activated, the blaster produces its characteristic ballistic response: it provides a non-specific, short-lived burst of transient attentional enhancement (TAE) to late layers in *both* key and response pathways. This is mediated by the connections marked (b) and (c) in Figure 1a. This boost provides the maximally active key and response types enough activation to initiate a combined binding process in the binding pool.

Response Pathway All response features, including that of the target, are treated equally in the response pathway.



(a) Architecture



(b) Dynamics

Figure 1: The 2f-ST^2 model

At the response Type layer, distractors presented in temporally proximal (i.e., -2, -1, +1 and +2) positions relative to the target can be co-active with the target response type. In order to generate a binding, the conflict between co-active response type nodes is resolved by weak lateral inhibition. Thus, only strongly active response types can initiate feature binding. Depending on relative timing of blaster firing and strengths of co-active response types, the target response type or a temporally proximal distractor can win this competition. This competitive interaction produces different possible binding outcomes, and thus the possibility of conjunction errors.

The Binding Pool The binding pool consists of two disjoint node subsets. The first subset maintains associations between key types and tokens. It receives activation from key type nodes via the connection labelled (d) in Figure 1a. Similarly, the second subset maintains associations between re-

sponse types and tokens. Thus, a successful binding involves two binding pool nodes, one in each subset. Activation from the two pathways drives the corresponding nodes in the binding pool subsets above threshold. These nodes consequently associate one token with two type nodes, one in the key pathway and the other in the response pathway.

Tokens The 2f-ST^2 model uses tokens to store episodic, temporally ordered representations of targets. During encoding, the currently active token inhibits all binding pool nodes of other tokens via the connection labelled (g) in Figure 1a. During binding, a token receives activation from a pair of binding pool nodes, via the connection labelled (h) in Figure 1a. As previously discussed, these binding pool nodes are in turn excited by a pair of (key-response) type nodes. The successful completion of binding yields an association between the active token, a key type and a response type, through sustained activation in the binding pool.

Summary In a trial, the relative time taken to process key and response features of items determines which response feature benefits most from the blaster's enhancement. Thus, it determines the bindings produced. Across simulated trials encompassing a variety of feature strengths, the model produces a range of binding outcomes, making up a response distribution (see bottom of Figure 1b). In most trials with the default model configuration (see below), the target response feature is the maximally active node in its pathway at the time of blaster firing. Hence, it wins the competition, and is successfully bound into a token, thereby producing a correct report. On some trials, response features of items before (-1 or -2 positions) or after (+1 or +2) the target are more active when the blaster fires, and get encoded instead. Such errors occur because of variation in the strength of key and response features, in addition to variation in processing delay in the pathways (see below). Of course, the relative probability and temporal position of conjunction errors depends on the model's specific configuration.

Configuration For all simulations, model configuration and parameters are kept unchanged (cf. Computational Methods section), except for the following three.

τ_D is a random, *varying* gaussian (mean = 0; s.d. = 15ms) delay. It is repeatedly sampled, once per item in the stream. For a particular RSVP item, τ_D is the same in the key and response pathways, but the bottom-up input strengths in these pathways vary independently. Consequently, the parallel processing of these features in Stage 1 is not perfectly synchronous. τ_D introduces temporal noise, generating conjunction errors and broadening response distributions.

τ_K and τ_R add a *fixed* positive or negative delay to the processing of all features in the key and response pathway, respectively. They are fixed within a complete simulation run. Either τ_K or τ_R is varied across a pair of runs, to model a key or a response feature manipulation. In the 'default' configuration, both τ_K and τ_R are set to 0. To simulate an isolated change in processing time within a pathway, τ_K or τ_R is adjusted.

Behavioural Predictions of the 2f-ST^2 Model

Isolated behavioural manipulations of the key and response pathways have been investigated by Botella et al. (2001). These manipulations change the correct report rate, and shift the loci of response distributions. Such shifts can be quantified using metrics like the *Average Position of Intrusions* (API) (Botella et al., 2001).

Key Feature Manipulation According to the 2f-ST^2 model, increasing key pathway processing delay will increase the proportion of post-target errors. This is because such an increase causes activation to reach the key Type layer later, in turn delaying blaster firing to a time at which post-target response items are most active. They consequently benefit more from its enhancement and get bound.

Exp. 1A from (Botella et al., 2001) provides an isolated manipulation of key feature processing time. Participants viewed an RSVP sequence of coloured words, and were asked to identify the colour of the only animal word (see Figure 2A). The frequency of the word in language use was manipulated: high-frequency (HF) words corresponded to key features that could be processed quicker than low-frequency (LF) words. Accordingly, a significant post-target shift and later API was found for LF compared to HF words (see top panel in Figure 2C). In addition, correct reports reduced from the HF to LF conditions.

Figure 2B illustrates how the 2f-ST^2 model simulates this data. The τ_K parameter is set to 0ms and 40ms for the HF and LF conditions, respectively, reflecting increased key feature processing delay in the latter. This manipulation delays average blaster firing latency for the LF condition, increasing the probability that it enhances post-target response features. The lower panel in Figure 2C shows the resulting post-target shift, and is mirrored by the corresponding simulated API values of -0.11 and 0.78 for the HF and LF conditions, respectively. In addition, the model generates more correct reports in the HF condition (62% vs. 55%). This post-target shift and reduction in correct reports replicate Exp. 1A from (Botella et al., 2001).

The BBS model also predicts a post-target shift with an isolated increase in mean key feature processing time. However, an important distinction between the two models relates to correct reports. The BBS model predicts that correct reports will *always* decrease along with an API increase. In contrast, 2f-ST^2 does not mandate a reduction in correct reports alongside a post-target shift. Whether the number of correct reports increases, decreases or remains unchanged along with a post-target shift depends on the pair of conditions being compared. Shortly, we present new behavioural data supporting 2f-ST^2 in this respect.

Response Feature Manipulation An isolated increase in response pathway processing delay means that, when the target key feature reaches the key type layer, pre-target response features are more likely to be active at the response type layer. Consequently, they benefit more often from the blaster's enhancement, increasing their likelihood of binding to a token,

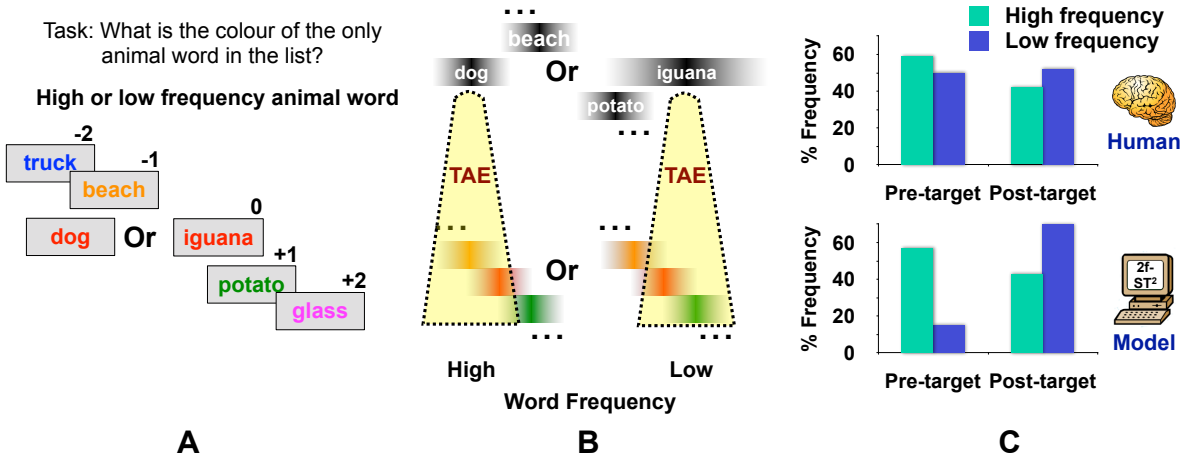


Figure 2: Key feature manipulation in (Botella et al., 2001) simulated by the 2f-ST^2 model.

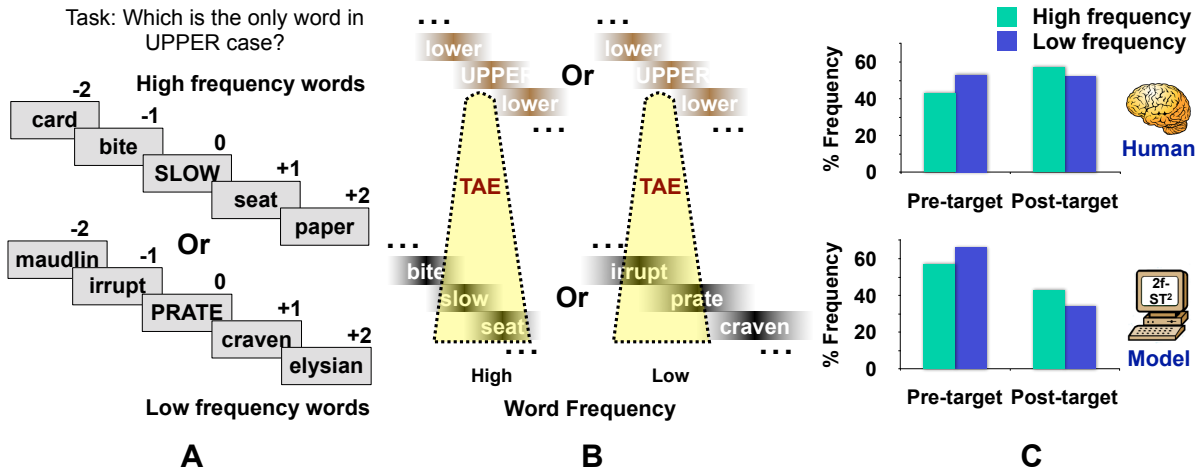


Figure 3: Response feature manipulation in (Botella et al., 2001) simulated by the 2f-ST^2 model.

producing a pre-ward shift in the response distribution.

Exp. 2 from (Botella et al., 2001) describes an isolated response feature manipulation, where target words were presented in uppercase, in a stream of lowercase distractor words (see Figure 3A). Response feature processing delay was manipulated using LF or HF words. Presumably, since LF words produced an increase in response pathway processing delay, a pre-target (HF to LF) shift was found (Figure 3C upper panel). Accordingly, LF words had a more negative API. However, there was no significant change in the number of correct reports.

To simulate this data in 2f-ST^2, τ_R (the fixed response pathway processing delay) is set higher (10ms) for LF than for HF words (0ms). Thus, LF item response features reach the Type layer later, and pre-target distractors have highest activation when the blaster fires (see Figure 3B), producing a more negative API for LF words. 2f-ST^2 replicates this pattern (Figure 3C lower panel), simulating API values of -0.11 and -0.31 for the HF and LF conditions, respectively. In addition, correct reports are the same (62%) in the two conditions. Thus, 2f-ST^2 replicates the findings in Exp. 2 from (Botella et al., 2001).

Again, the 2f-ST^2 and BBS models simulate response feature manipulations similarly. However, the BBS model predicts that correct reports are unchanged when the response pathway is manipulated. This is because the relative proportion of trials processed by the AF or SG routes remains unchanged. In contrast, 2f-ST^2 predicts that sufficiently large response pathway manipulations will change correct report rates.

Reaction Times for Response Positions We show that 2f-ST^2 replicates existing RT data (Botella, 1992) using a single binding mechanism, weakening the need for the BBS model's dual-route approach. Botella (1992) showed that correct re-

port RTs are fastest, followed by pre- and then post-target errors (see Figure 4). This is a counter-intuitive finding, as at first sight, one would expect pre-target errors to be earliest, followed by correct reports and then post-target errors. Furthermore, one might expect that a parallel model like 2f-ST^2 would generate this pattern. Consequently, this RT data is a crucial test of 2f-ST^2.

In 2f-ST^2, RTs are modelled by the time at which the token completes binding to a pair of key and response type nodes. This event is indexed by the trace neuron (Bowman & Wyble, 2007) of the active token reaching 75% of its maximal activation. To generate RTs, trials are grouped by response position and separately averaged. As seen in Figure 4, 2f-ST^2 replicates the qualitative pattern of data reported by Botella (1992).

2f-ST^2 produces this pattern because, for correct reports, the blaster fires temporally close to the activation peak of a strongly active target response type node. In this situation, this node quickly wins the competition with co-active response type nodes. In contrast, conjunction errors arise when the target's key and response features are relatively weak. Because of a weak target key feature, the blaster fires later. When a pre-target item's response feature is relatively stronger, the corresponding type node benefits more from the blaster, wins the competition and produces a pre-target error. A corresponding pattern arises for post-target errors. For pre- and post-target errors, the blaster fires temporally further from the peak activation of the response type that gets bound. Hence, these types take longer to overcome the response type layer's lateral inhibition. This increases mean RTs for conjunction errors. Naturally, post-target error means are later than pre-target error means, as the blaster fires later.

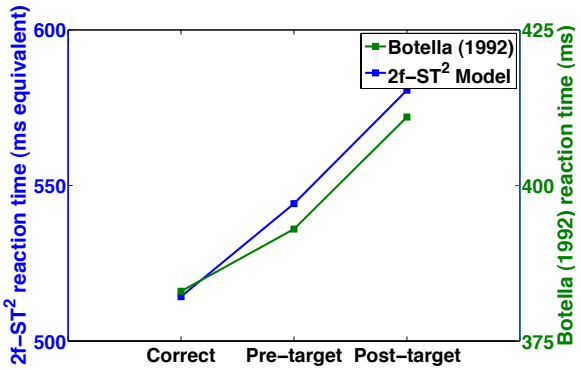


Figure 4: RT data from (Botella, 1992) explained by 2f-ST^2.

The Temporal Feature Binding Experiment

We present new data that refutes a prediction from the BBS model. We presented coloured letter or symbol targets within RSVP streams of digit distractors (SOA 94ms). The key feature was target category (letter or symbol), and the response feature was colour (cf. Experimental Methods). Figure 5A

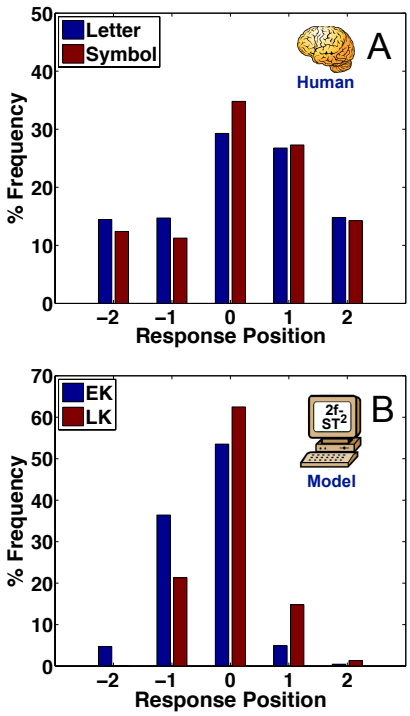


Figure 5: Simulation of new behavioural data by 2f-ST^2.

depicts the resulting response distributions. The APIs for letter (0.18) and symbol (0.31) targets are significantly different ($F(1,13) = 9.7$, MSE = 0.01, $p < 0.01$). Hence, this experiment manipulates key feature processing delay. In addition, correct report rates vary across the conditions: 29% for letter and 35% for symbols, with a marginally significant difference ($F(1,13) = 3.24$, MSE = 65.9, $p = 0.09$). This marginal *increase* in correct reports accompanying the letter to symbol post-target shift contradicts the BBS model, which mandates a *decrease* in correct reports.

To simulate this with 2f-ST^2, we manipulate key pathway delay, producing the *early key* (EK) and *late key* (LK) conditions, with τ_K values of -40ms and 0ms, respectively (see Figure 5B). Going from the EK to the LK condition, API increases from -0.86 to -0.11, and correct reports increase from 54% to 62%. That is, the post-target shift is associated with an increase in correct reports. Hence, the pattern of changes between the EK and LK conditions are qualitatively equivalent to those observed in the human data, although the increase in correct reports did not quite reach significance.

Conclusions

The 2f-ST^2 model provides a detailed neural explanation of attentional enhancement in temporal feature binding. The theoretical framework underlying this explanation (Bowman & Wyble, 2007) already elucidates a broad profile of temporal cognitive phenomena, including the attentional blink, repetition blindness, episodic perception, etc. This level of description in 2f-ST^2 extends beyond that afforded by the

BBS model. The latter leaves unspecified neural realisation of the AF and SG routes, and indeed, the mechanistic basis for choosing between them. In comparison, 2f-ST^2's implementation forces us to elaborate on how cognitive processes might be neurally realised. It explains how a single neural construct can produce different behavioural outcomes. Importantly, the process of temporal feature binding in 2f-ST^2 does not have any knowledge of the 'correct' response feature. Rather, correct responses and conjunction errors are all generated by the same underlying architecture. Importantly, this architecture explains counter-intuitive RT data (Botella, 1992), which motivated the dual-route approach of the BBS model. In this respect, 2f-ST^2 harks back to earlier, informal, parallel models of selective attention (Keele & Neill, 1978), which were seen as incompatible with the RT data. Thus, we argue that 2f-ST^2 is more parsimonious in mechanism, and rejuvenates the adage that correct reports are indeed nothing more than fortunate conjunctions.

Computational Methods

2f-ST^2 extends the ST^2 model (Bowman & Wyble, 2007), with the minor parameter modifications described below.

Stage 1 ST^2's Stage 1 was replicated to generate 2f-ST^2's key and response pathways. The following changes were made to 2f-ST^2's key pathway, in comparison to ST^2's Stage 1. Firstly, to decrease Type layer activation strengths, the Item to Type layer weights were reduced from 0.015 to 0.013. In addition, the Type layer leak current was increased from 0.07 to 0.0715. Secondly, to ensure that Type layer on neurons were suppressed at the correct times, weights from Type layer off neurons to on neurons were decreased from -0.12 to -0.15. Finally, to activate the blaster more strongly from the target type node, the Type layer to blaster Input weight was increased from 0.02003 to 0.02803.

Leak current and connection weight settings for the layers in Stage 1 of the response pathway are presented in Tables 2 and 3 of Chennu (2009). Weak lateral inhibition was added between Type layer on neurons (connection weight -0.05) in the response pathway. In addition, Item and Type layers did not influence the blaster, but were enhanced by it. Task demand was absent at the response pathway Type layer.

Stage 2 The Stage 2 binding pool was expanded to represent associations between response pathway types and tokens (one per combination of 4 tokens and 25 response types). To compensate for the altered activation dynamics at the token level, Stage 2 weights were adjusted as in Table 3 of Chennu (2009).

Dynamics Across simulation runs, target key feature strength was varied across a range, while distractor strengths remained constant. However, to simulate the generation of correct reports and conjunction errors with feasible simulation times, response feature strengths of the target and proximal distractors (positions -1, -2, +1 and +2) were varied. Other distractor response feature strengths were constant across all trials. For further details see Chennu (2009).

General Configuration A delay τ_D was added to the processing of all features in both pathways. Within each run, τ_D was sampled once per item in the stream, from a gaussian distribution (mean = 0; s.d. = 15ms). Key/response pathway processing time manipulations involved adding a constant delay (τ_K/τ_R) to feature processing in the relevant pathway. In the default configuration, τ_K and τ_R were set to 0 (cf. (Chennu, 2009)).

Key and Response Feature Manipulations To simulate the HF condition in Exp. 1A from Botella et al. (2001), 2f-ST^2 was run in its default configuration (τ_K = 0ms; τ_R = 0ms). The LF condition was simulated by introducing a fixed delay of 40ms (τ_K = 40ms; τ_R = 0ms) in key feature processing. Similarly, the HF condition of Exp. 2 in Botella et al. (2001) was simulated by running 2f-ST^2 in its default configuration, and the LF condition by introducing a fixed delay of 10ms (τ_K = 0ms; τ_R = 10ms) in response feature processing.

Experimental Methods

Participants 14 students undertook the experiment. They were free from neurological disorders and had normal or corrected-to-normal vision. The study was approved by the local ethics committee.

Stimuli and Apparatus We presented dark grey alphanumeric characters and symbols surrounded by coloured squares. The characters/symbols and coloured squares subtended maximal visual angles of 1.2° x 1.2° and 1.36° x 1.36°, respectively.

Procedure Participants viewed two blocks of 180 trials each, which began with a task instruction indicating whether a letter or a symbol would be the target. A trial began with a central, white fixation cross, which after 500ms, turned into a white arrow indicating the side to monitor for targets. After 200ms, two 16-item streams were presented at equal distances (2.7° visual angle) to left and right of fixation, at a rate of 94ms per item with no inter-stimulus interval. Colours for the squares surrounding items were sampled from red, green, blue, yellow and cyan. Distractors were digits randomly sampled from 2, 3, 4, 5, 6, 7, 8, 9. The target was presented on one side of fixation, at a random position between 6 and 12 in the stream. Targets were letters or symbols (depending on the block), randomly chosen from D, E, G, K, L, T, U, V or #, %, £, ¥, ≪, ×, Þ, ÷. Colours for the target and proximal distractors (positions -1, -2, +1, +2) were assigned using a random permutation of the above five colours. Participants were told to direct their covert attention towards the stream indicated by the arrow, search for the target item, note both its identity, and the colour of the surrounding square. At trial end, they were asked to select the identity of the target item from a 6-option menu. The first 5 consisted of the target plus 4 other items, randomly chosen from the block's target set. The menu's final option was 'None of the above'. A further response menu required selection of the colour of the square surrounding the target, and again contained 6 options. The first 5 were a random permutation of the 5 colours used in the experiment, and the last option was 'None of the above'.

Computational Modelling The *early key* (EK) condition from the temporal feature binding experiment was simulated by introducing a key feature processing delay of -40ms (τ_K = -40ms; τ_R = 0ms). The *late key* (LK) condition was simulated by running 2f-ST^2 in its default configuration (τ_K = 0ms; τ_R = 0ms).

References

Botella, J. (1992). Target-specified and target-categorized conditions in RSVP tasks as reflected by detection time. *Bulletin of the Psychonomic Society*, *30*(3), 197-200.

Botella, J., Barriopedro, M. I., & Suero, M. (2001). A Model of the Formation of Illusory Conjunctions in the Time Domain. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(6), 1452-1467.

Bowman, H., & Wyble, B. (2007). The Simultaneous Type, Serial Token Model of Temporal Attention and Working Memory. *Psychological Review*, *114*(1), 38-70.

Chennu, S. (2009). *The temporal spotlight of attention: computatioanl and electrophysiological explorations*. Unpublished doctoral dissertation, School of Computing, University of Kent, Canterbury CT2 7NF, UK. Available from `http://www.cs.kent.ac.uk/pubs/2009/3054`

Keele, S. W., & Neill, W. T. (1978). Mechanisms of attention. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of Perception* (Vol. 9, p. 3-47). New York: Academic Press.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171-178.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97-136.