Temporal Selection in Dynamic Displays:

Sensory Information Persists Despite Masking

Charles J. H. Ludowici

School of Psychology, Faculty of Science

The University of Sydney

2019

A thesis submitted to fulfil requirements for the degree of Doctor of Philosophy

Table of Contents

Statement of Originality and Ethics Approval

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

The research conducted in this thesis was approved by the Human Research Ethics Committee of The University of Sydney (protocol number: 2016/800).

Charles J. H. Ludowici

Acknowledgements

The work presented in this thesis would not have been possible without support from many people, all of whom deserve thanks. Alex Holcombe's deep knowledge of human vision and cognition, dogged pursuit of rigour, and critical thoughts on theory, experimental design and writing made for invaluable supervision. Anna Doherty supported me without question, is directly responsible for my postdoc, and must be pretty relieved the whole thing is over. My parents provided more moral (and financial) support than I probably deserved. My labmates - Will Ngiam, Ryohei Nakayama, Kim Ransley and Chris Bush - were a vital source of scientific and not-so-scientific conversation. The same goes for the postgrad and postdoc community at USyd (in no particular order): Shuier Han, Indako Clarke, Justine Greenaway, Julie Chow, Huihui Zhang, and Nahian Chowdhury. Jeremy Wolfe was kind enough to host me in his lab in Boston in the early part of 2019 for a very stimulating visit. Vincent Wozniak O'Connor spent many weekends driving deep into NSW to walk and ride bikes. Jason L'Ecuyer and Ben Corio provided pleasant company on some of those trips. Dan Bishton coaxed me out of my cave when I needed it.

"*Rather the long doubts: That this labor tears up the mind's moorings; that, though life may be important in the scheme, awareness is an imperfect tool with which to face it.*"

S. R. Delany, Dhalgren, 1974

Abstract

The visual system receives a dynamic stream of information, but it has a limited capacity and must deploy its resources to behaviourally relevant stimuli - a process referred to as "attention". Rapid serial visual presentation (RSVP) is an experimental method for investigating attention's time course by presenting a rapid sequence of stimuli at a single location. Attentional selection in both naturalistic viewing and RSVP is limited by masking, and many models of selection in RSVP assume that masking terminates sensory memory for stimuli that are no longer present. However, there is indirect evidence that information about unselected RSVP stimuli may persist in a buffer despite masking. In this thesis we directly investigate buffering and selection of a cued item from one of multiple simultaneous RSVP streams. We use mixture modelling to analyse reports from only those trials in which participants identified a letter in response to the cue, and outline a novel quantitative test for buffering (Chapter 2). This provides new insights into the temporal variability of selection with exogenous and endogenous cues (Chapter 3). A series of experiments show that participants can select buffered representations, despite masking, and this appears to be related to the number of simultaneous RSVP streams (Chapter 4). We also investigate possible contributions of crowding and eccentricity to selection (Chapter 5). RSVP provides a measure of attention's timing that replicates classic attentional effects. However, participants appear to dedicate attention to the streams prior to the cue's appearance. When there are few streams, this leads to attentional speeds fast enough to select a stimulus representation that persists briefly, despite the masking inherent in RSVP. This falsifies theoretical claims about masking in RSVP, and demonstrates that the dynamic nature of naturalistic viewing does not prevent selection from sensory memory.

Chapter 1: Introduction

1.1 Capacity limits in human vision

Human visual processing is capacity limited, and this capacity is exceeded by the stimuli present in many everyday scenes. To circumvent this capacity limit, the visual system allocates its resources flexibly in order to process behaviourally relevant stimuli - a phenomenon known as “attention”. The way in which attentional resources are allocated and the time course of this allocation provide constraints on the ability to respond to a stimulus that may be presented only briefly at a particular location, but is important to identify. For instance, a driver on a busy road needs to attend and respond to a child about to walk onto a pedestrian crossing, but need not allocate their limited attentional resources if the child is, say, on the footpath nearby. Here, we are interested in the conditions under which attentional selection from a busy display occurs and how the time course of selection changes with the amount and kind of information presented to an observer.

The retina has a massively parallel architecture, but it feeds information to various capacity limited processes. Even visual performance based on information that is thought to be well represented at the early stages of cortical visual processing, such as orientation judgements (Lennie & Movshon, 2005), suffer from limited capacity (Lavie, Beck, & Konstantinou, 2014). Simple detection tasks rely on limited capacity. Response times reveal a cost associated with detecting a change in the location of two dots relative to a single dot (Hawkins, Houpt, Eidels, & Townsend, 2016). Accuracy for identifying the features in a visual display also suffers when those features are split across objects, rather than part of a single object (Duncan, 1993). Memory processes demonstrate capacity limits as well. Visual working memory has famously limited capacity (Luck & Vogel, 1997), as does sensory memory. When asked to report all the elements in a briefly presented array of stimuli, participants are poor, but they can select a subset for report with a much higher accuracy than the whole-report predicts (Averbach & Coriell, 1961; Sligte, Scholte, & Lamme, 2008; Sperling, 1960).

Binding the visual features of an object into a veridical representation is poor when participants must simultaneously complete a demanding task. Under these conditions participants may report illusory conjunctions - pairings of objects and features that did not occur together (Treisman & Schmidt, 1982). These errors demonstrate that the process of binding is disrupted when resources are directed elsewhere, or stimuli are presented too quickly to recruit them. Such errors occur because the system responsible for binding operates on a limited capacity.

When searching for items that are conjunctions of features that must be bound together, people classically exhibit a lengthening in response time as the number of items in the display increases (but see Nordfang & Wolfe, 2014). This cost is typically greater than that found in searches for targets defined by a single feature (Treisman & Gelade, 1980; Wolfe, 1998). The conjunction search cost is thought to either represent a capacity limit for processing one item at a time (Treisman & Gelade, 1980; Wolfe & Gray, 2007), or the dilution of a limited capacity across the multiple stimuli (Algom, Eidels, Hawkins, Jefferson, & Townsend, 2015; McElree & Carrasco, 1999).

Neurally, simultaneously presented stimuli result in less activation for each item than stimuli presented on their own. When participants are shown irrelevant peripheral stimuli while they engage in a task at fixation, the activation in visual areas associated with the irrelevant stimuli decreases as the number of irrelevant stimuli increase, such that each stimulus is associated with weaker signal (Beck & Kastner, 2005, 2007; Kastner, de Weerd, Desimone, & Ungerleider, 1998; Kastner et al., 2001). The interference is apparent in later visual areas associated with object representations (V4 and the inferior temporal cortex) relative to earlier visual processing areas, where it is often not observed. These suppressive interactions between stimuli indicate that representations compete for activation, consistent with the idea of a limited resource (Beck & Kastner, 2009; Desimone & Duncan, 1995). Similar suppressive interactions between competing stimuli within cortical receptive fields were observed in macaques using single-cell recordings by Reynolds, Chelazzi and Desimone (1999).

These neural and behavioural results indicate that the visual system is defined by limited capacity. When this capacity is engaged elsewhere, there is a degradation in visual processing as assessed by the accuracy of behavioural responses, as indicated by illusory conjunctions and the efficiency of visual search for conjunction targets. The capacity of the visual system may be taxed or exceeded by the simultaneous presentation of multiple stimuli. This leads to behavioural costs, such as the slowing down of accurate responses in visual search for conjunction targets. It also leads to neural competition, simultaneously presented stimuli elicit weaker activation in the human visual cortex and visual cortical cells of primates.

1.2 Attention

Given the limited capacity of visual processing, how does the visual system deal with the stream of visual information it receives? Naturalistic visual scenes are dynamic, changing due to the actions of an observer and movement within a scene. There may be a lot of action in a scene, but given limited processing capacity, a human observer must dedicate visual processing resources selectively to the aspects of the scene most relevant for their behaviour. Furthermore, the aspects of a scene that are relevant to behaviour are not static, so the allocation of resources must be dynamic. The visual system’s ability to achieve this is referred to as “attention”.

Attention is typically associated with the central part of the visual field, which is advantaged relative to the periphery in many aspects of visual processing. It has higher spatial acuity (Weymouth, 1958), different retinal cell topography (Curcio, Sloan, Kalina, & Hendrickson, 1990) and better form vision (Bouma, 1970; Strasburger, Rentschler, & Jüttner, 2011). Attentional processing is thus often associated with an observer's fixation location, and changes in the location of fixation brought about by eye movements are interpreted as "overt" attention shifts. However, attention can also be allocated to locations in the visual field that differ from central vision. These "covert" shifts of attention, described as such because they cannot be inferred from eye movements, are of particular interest to vision scientists. They indicate the flexible allocation of visual resources that need not be related to changes in the retinal signal, but are still associated with changes in visual performance. They also seem to operate on a timescale that is faster than that needed for an observer to program and execute an overt attention shift, otherwise known as a saccade. Saccades occur typically within 250 ms of the appearance of a saccade target (Saslow, 1967), but covert attention can yield peak performance within 100 ms of a cue (Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989). In this thesis I am concerned with these covert attention shifts.

Attention is typically studied as a phenomenon in which a particular region of visual space is attended to and thus prioritised in the absence of eye movements to that location. Attention is manipulated by cues providing information about a spatial location (Posner, 1980). Such cueing typically leads to an advantage in processing for information in the cued region, as shown by psychophysical measures like target detection reaction times or contrast thresholds. The idea that visual processing can prioritise particular sources of visual information over others is as old as experimental psychology itself. Wundt spoke of attended objects as characterised by "sharp discrimination" from other, unattended, objects (Wundt, 1897, p. 209) and investigated the temporal qualities of attention (Carlson, Hogendoorn, &

Verstraten, 2006). Helmholtz described the phenomenology of attending to different locations of the visual field and how doing so advantaged information from that region over others (Yantis, 1998).

Sperling's (1960) classic investigation of sensory memory was an early experimental demonstration of how visual resources could be directed in a way that advantaged a subset of information in a visual array. He presented subjects with arrays of 6 - 12 letters arranged in 2 - 3 rows of equal length. Presentation of these stimuli was brief, 50 ms, and in separate experiments participants were instructed to report the whole array or a single row, indicated by a tone presented after the array was terminated. Report accuracy for the whole array was poor - an average of 4.3 letters. When participants were instructed to report only part of the array, their accuracy was much higher, suggesting that participants had on average about 9 letters available to them at the time of the cue. The accuracy of the partial report condition was negatively related to the length of time between the offset of the array and the tone. A similar phenomenon was described, independent of Sperling, by Averbach and Coriell (1961) using a visual stimulus to indicate the reported subset rather than a tone.

Sperling's (1960) work, along with Averbach and Coriell (1961), demonstrated an attentional effect. In the whole report condition, the process of recalling the entire array took longer than the persistence of the visual information, so that only some proportion of the array could be reported before the information was no longer accessible to the observer. In the partial report conditions, participants could direct attention to a subset of the persistent information and maintain it while the rest of the information about the array decayed. While the authors of these studies were mainly interested in the form of memory underlying the partial report advantage, the paradigm provided early experimental evidence for an attentional effect. Visual resources were dedicated to a particular subset of the information available to the visual system, selecting and maintaining this information while other, unselected, information decayed.

Much of the following work on attention investigated its ability to select information based on spatial location. Eriksen and Hoffman (1972; 1973) conducted a series of experiments in which they presented participants with letters arrayed about fixation, cued the location of a letter to be named by the participant and varied the time between the onset of the cue and the onset of the letters. They found that reaction times for identifying the letter decreased as the lag between the cue and the letters increased (Eriksen & Hoffman, 1973). When the cue and the target letter were simultaneous but there was a lag between those stimuli and distracting letters, reaction times decreased as a function of lag (Eriksen & Hoffman, 1972). Posner, Snyder and Davison (1980) presented subjects with a central arrow that was informative about the location of a target on the majority of trials (valid trials) or uninformative (invalid trials). The cue provided no information about the correct response. Participants also saw trials in which the cue did not identify a particular location (neutral trials). Responses for target identification were faster in the valid trials than the neutral trials, and slower in the invalid trials than neutral trials (See also Posner, 1980). Thus attention can be applied to different locations in the visual field. When its location coincides with that of a target stimulus, responses are more efficient. When it is directed away from a target stimulus, responses are less efficient. Critically for our purposes, the Eriksen and Hoffman (1972; 1973) results suggest a temporal component to attention - it takes some period of time after the onset of a cue for attention to select a particular region of the visual field.

1.2.1 Exogenous and Endogenous Attention

Nakayama and Mackeben (1989) and Müller and Rabbitt (1989) described two components of visual attention, now commonly referred to as exogenous and endogenous attention (Carrasco, 2011; Posner, 1980). These forms of attention reflect automatic and voluntary orienting of attention, respectively, and they operate at different timescales in response to different kinds of cues.

Nakayama and Mackeben (1989) and Müller and Rabbitt (1989) set out an influential method for investigating exogenous attention. In these experiments, participants maintained fixation and either saw a 50 ms brightening of one of four peripheral boxes to indicate which of the boxes contained a target stimulus (Müller & Rabbitt, 1989) or the onset of a square indicating the location of a visual search target, followed by the search target and distractors (Nakayama & Mackeben, 1989). Exogenous attention is typically investigated using rapid-onset peripheral cues (Carrasco, Giordano, & McElree, 2006; Carrasco & McElree, 2001; Carrasco & Yeshurun, 1998; Cheal & Lyon, 1991; Giordano, McElree, & Carrasco, 2009; Hein, Rolke, & Ulrich, 2006). By varying the period of time between the cue's onset and the critical stimulus on each trial and measuring participants' accuracy, the authors assessed the amount of time needed for attention to arrive at the cued location.

In both Nakayama and Mackeben (1989) and Müller and Rabbitt (1989), accuracy in response to the peripheral, exogenous cue increased with cue-target lag to a peak at lags of approximately 100-120 ms. After this peak, accuracy declined, indicating that the cueing advantage is transient in these circumstances. Peak accuracy of 100 - 120 ms is commonly seen with peripheral cues (Carrasco, 2011; Folk, Remington, & Johnston, 1992; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Posner & Cohen, 1984; Remington, Johnston, & Yantis, 1992). Importantly, the allocation of attention associated with a peripheral cue does not appear to be under the observer's control. The same changes in accuracy in response to a peripheral cue can be observed when the participant knows the target's location - either because it does not change (Nakayama & Mackeben, 1989) or an earlier symbolic cue indicated the location (Müller & Rabbitt, 1989). When participants know that the cue never indicates the target location - information that should produce attempts to ignore it - it still produces a reaction time increase, suggesting that attention oriented to the location of the uninformative cue (Remington et al., 1992). Likewise, if a peripheral cue only rarely indicates the target's location, such that participants should have little incentive to attend to it if this attention is under their control, there are costs for cueing a location other than that of the target and benefits when the target and cue occur at the same location (Folk et al., 1992). Finally, manipulating the proportion of trials in which an exogenous cue indicates the target stimulus does not appear to affect the cueing benefit (Giordano et al., 2009).

It is important to note that the peak at cue-target lags of 100 to 120ms does not mean that on average, attention arrives at the cued location at this time after the cue on each trial with this lag. The relationship between accuracy and lag using this method instead likely describes the cumulative distribution of attention's arrival times from each set of cue-target lag trials. An accuracy advantage at a 100 ms lag trial may mean that attention arrived at the cued location at or before 100 ms after the cue's onset. In studies using cue-target lags to measure attention's time course (i.e. Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989) a number of lags are used, but this method does not provide a measure of attentional arrival times on each trial. Judgements about the timing of exogenous attention are also complicated by the decline in accuracy after the peak. For a given cue-target lag after the peak, it is not clear when attention arrived at the cued location or when it disengaged. Other methods - those used in this thesis and described below - may better measure the timing of attention.

Endogenous attention is slower than exogenous attention, it can be sustained at a particular location and reflects a process that is under the observer's control. Nakayama and Mackeben (1989) demonstrated that participants could endogenously apply their attention to a particular location and hold it there, resulting in a boost in accuracy at that location. They used a cue, present for the entire experiment at this location, or informed participants that the target, when present, would always be at the same location. Other studies manipulated endogenous attention with a symbolic cue indicating the likely location of a target stimulus, after Posner (1980). Endogenous attention is thought to be voluntary because symbolic cues require interpretation, unlike exogenous cues which occur at the cued location. Müller and Rabbitt (1989) manipulated endogenous attention with such a symbolic cue. They presented an arrow at fixation for 50ms that pointed towards the target location on 50% of trials, which were termed the valid trials. On valid trials, accuracy increased over cue-target lag, peaking at around 275-300ms. This increase was thus much slower than with the peripheral, exogenous cue. Such a time course is typical of endogenous attention (Cheal & Lyon, 1991; Liu, Stevens, & Carrasco, 2007; Müller & Rabbitt, 1989).

Endogenous attention is under the observer's control. Manipulations of cue validity affect the benefits associated with endogenous cueing. That is, participants appear to learn about the probability that the cue indicates the location of the target and allocate their attention based on this belief. Endogenous cues that indicate the location of a target stimulus on only a small proportion of trials produce a smaller benefit, when valid, than cues that are more likely to indicate the target's location (Eriksen & Yeh, 1985; Giordano et al., 2009).

The effects of cued attention brought about by endogenous and exogenous cues reflect an important aspect of flexible human visual performance (for a review see Carrasco, 2011). Cueing attention can result in visual processing that excludes the effects of external luminance noise on orientation discrimination (Dosher & Lu, 2000b, 2000a). Cues may improve the representation of the cued stimulus, as reflected in by contrast sensitivity (Carrasco, Penpeci-Talgar, & Eckstein, 2000; Foley & Schwarz, 1998; Huang & Dobkins, 2005; Lu & Dosher, 1998, 2000; Solomon, Lavie, & Morgan, 1997). Spatial resolution is improved by cueing (Carrasco, Williams, & Yeshurun, 2002; Yeshurun & Carrasco, 1998, 1999). Asynchronies in the binding of colour and motion (Moutoussis & Zeki, 1997) can be eliminated by an exogenous attention cue (Holcombe & Cavanagh, 2008).

It is thus well-documented that attentional cueing improves performance, consistent with the allocation of visual processing resources. More evidence for this claim comes from observations that attentional cueing can attenuate the effects of visual capacity limits. Two such capacity limits are the set size effect in visual search and its interaction with the kind of search being performed.

Visual search reaction time increases with the number of items in a display - the set size effect. Treisman and Gelade's (1980) feature integration theory was based in part on their observation that set size effects on reaction time are only seen with conjunction searches. However, set size effects on accuracy do occur for feature search (Cameron, Eckstein, Tai, & Carrasco, 2004; Carrasco & McElree, 2001; Eckstein, 1998). This set size relationship is stronger for conjunction searches, where features must be bound together, relative to feature searches and this increased inefficiency is thought to reflect attentional recruitment in feature binding (Treisman & Gelade, 1980; Wolfe & Gray, 2007).

When the location of a visual search target is cued, the relationship between set size and reaction time is attenuated (Carrasco et al., 2006; Carrasco & McElree, 2001; Giordano et al., 2009). This occurs for exogenous cues (Carrasco et al., 2006; Carrasco & McElree, 2001; Giordano et al., 2009) as well as endogenous (Giordano et al., 2009). Critically, the effect is greater for conjunction searches than feature searches (Carrasco et al., 2006; Carrasco & McElree, 2001; Carrasco & Yeshurun, 1998), although the latter may only show set size effects in measures relating to accuracy (Carrasco et al., 2006; Carrasco & Yeshurun, 1998). This demonstrates that attentional cueing directs the visual system's limited resources in a way that eliminate the effects of those limits. Not only does cueing attenuate the effects of distracting information, but also it is most effective for conjunction searches which require more processing resources.

1.3 Temporal Selection in Vision: Measurement

The attentional cueing literature demonstrates that the visual system’s limited resources can be directed, voluntarily and involuntarily, to particular locations in the visual field. Doing so improves processing at those locations. The distinction between endogenous and exogenous attention has focused not only on the different levels of automaticity involved, but also on their different timescales. Endogenous attention requires a longer cue-target lag, on average, than exogenous attention. However, as discussed, cue-target lags do not provide a measure of attentional arrival times. At best, a set of trials showing a statistical advantage for an exogenous cue with, for example, a 100ms cue-target lag over an endogenous cue with the same lag tells us that exogenous attention was more likely to arrive at the cued location *at or before* 100ms. Therefore, the relationships between cue-target lag and accuracy, such as those described by Müller and Rabbit (1989) and Nakayama and Mackeben (1989), likely represent cumulative distributions of attentional arrival times. We can infer that, for any given cue-target lag, attention arrived at the cued location at or before that lag. Testing a range of cue-target lags lets us plot this cumulative distribution.

This is true, at least, for endogenous attention, which can be sustained at the cued location. However, exogenous attention has a transient component, at least in certain circumstances. Exogenously cued performance rises rapidly, but declines after a peak (Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Posner & Cohen, 1984). This represents attentional disengagement after cueing and it complicates the interpretation of the relationship between cue-lag and accuracy. For cue-lags that are longer than the most efficacious cue-lag, we do not know when attention arrived at the cued location and when it disengaged. The accuracy benefits associated with attentional cueing are only observable if attention can *deploy and sustain* at the cued location over a tested lag.

In order to investigate the timing of visual attention in a way that gives a more informative estimate of its arrival time, researchers have used rapid serial visual presentation (RSVP). In RSVP, a sequence of stimuli are shown sequentially at one spatial position in the visual field (Potter & Levy, 1969). RSVP was initially developed as a tool to investigate comprehension in reading (Forster, 1970; Potter, Kroll, & Harris, 1981) and memory processes in sequences of images (Potter, 1976; Potter & Levy, 1969), but it provides a valuable tool for investigating temporal phenomenon in visual attention. Participants are instructed to report a target stimulus from a sequence, or “stream”, of stimuli to assess the timing of attentional selection. This stimulus may be designated as the target by its appearance at the same time as a cue, or determined by some quality that makes it stand out (i.e. a red letter among grey letters). Given a rate of presentation that is fast enough - typically 8 - 12 items/second is used - participants will make errors in the form of reports of stimuli from the stream that were presented at times different from the cue. The period of time between the reported stimulus and the cue provides an estimate of the attentional selection latency on that trial.

RSVP is designed to replicate the visual world’s dynamic nature in a controlled setting (Potter & Levy, 1969). When we observe a scene, visual stimuli are replaced by new stimuli at the same visual locations when we make a saccade, move, or something in the scene moves. Stimuli presented close in time at the same location interfere with each other. Presenting a pattern at the same location after a target stimulus can result in a reduction in detection or accuracy for the target relative to when the target is presented without a subsequent stimulus - a phenomenon called “masking” (Enns & Di Lollo, 2000; Kahneman, 1968). Masking is thought to represent the termination or suppression of the target stimulus’s representation by the mask (Sperling, 1960), and it is a fundamental property of RSVP. In an RSVP stream, the sequential presentation of stimuli at the same spatial location means that each stimulus is masked by the stimulus following it, with the exception of the final stimulus in a stream.

The presence of masking in RSVP is thought to result in a report of the stimulus present when attention oriented to the stream in response to the target. Masking in RSVP is thought to prevent the selection of stimuli from sensory memory, which would reduce the apparent selection latency of attention by allowing it to select a stimulus that was no longer present. Many models of selection from RSVP assume that when a new stimulus is presented, the representation of any subsequent stimulus is terminated (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002) or that representations persist briefly despite masking, but are not selected (Olivers & Meeter, 2008; Wyble, Bowman, & Nieuwenstein, 2009). Others allow the persistence of only partial information about an object upon masking, such as its colour, and this is not available for attentional selection (Botella, Barriopedro, & Suero, 2001) or is not bound to an object representation (Vul & Rich, 2010).

In addition to investigating temporal selection under conditions of masking, RSVP allows us to assess how other factors of the visual display affect selection. For instance, there is evidence that temporal processing is better for stimuli that are further from fixation (Carrasco et al., 2006; Carrasco, McElree, Denisova, & Giordano, 2003; Hartmann, Lachenmayr, & Brettel, 1979; Tyler, 1987). By manipulating the eccentricity of the cueing stimulus and target, we can assess how the time course of attentional selection changes with eccentricity. We can also assess how the presence of multiple simultaneous streams of information affect selection. By changing the number of simultaneous streams, we can assess how increasing the number of potential locations of a cue affects attentional selection in response to it. RSVP thus allows us to mimic ecological concerns in temporal selection: masking, eccentricity and the number of potential locations of a cueing stimulus.

The serial position of an item that a participant reports from an RSVP stream relative to the cued item in that stream is the *Serial Position Error* (SPE) and this provides an estimate of attentional selection latency. The SPE measures the direction in time and latency of selection relative to the target on a particular trial. The target item on each trial has a serial position of zero. Selections of items after the target have positive SPEs. Those from before the target have negative SPEs.

Cue-target lags require comparisons in order to draw inferences about attention’s time course, and those inferences provide only cumulative information, as discussed. SPEs on the other hand, provide an estimate of the arrival time of attention on each trial. However this estimate is qualified somewhat by the quantisation inherent in most RSVP studies. Typically, RSVP streams are composed of discrete stimuli, each presented for an equal period of time (for an exception see Callahan-Flintoft, Holcombe, & Wyble, 2019). For example, if RSVP stimuli are presented at a rate of 12 Hz, each stimulus is presented for 83.3 ms . If attention is triggered by the onset of the cue and a participant reports the letter following the cued stimulus, all we can infer is that attention selected an item between 83.3ms and 166.6ms after the onset of the cue. That is, the discrete nature of RSVP stimuli provides a quantised estimate of attentional arrival times. This is not a direct measure of attention’s timing qualities, but it is better than the cumulative estimate provided by cue-target lag manipulations and gives a more readily interpretable distribution of attentional arrival times than such manipulations.

Furthermore, RSVP does not require attention to sustain at the cued location longer than is necessary to select a letter, unlike cue-target lags. In RSVP, attentional resources are no longer necessary at the cued location once the stimulus is selected, whereas in cue-target lag studies attention must deploy to the cued location and be sustained there until the target is presented.

1.4 Temporal Selection in Vision: Phenomena

Much of the use of RSVP to investigate temporal phenomena in vision has concentrated on the attentional blink (AB). The AB is observed when participants must select two targets presented at different times in a single RSVP stream. The classic finding is that successful selection of a first target from the sequence inhibits the selection of another target presented 200 - 500 ms afterwards (Broadbent & Broadbent, 1987; Raymond, Shapiro, & Arnell, 1992; Weichselgartner & Sperling, 1987). Most theorists believe that this so-called blink reflects the time course of attentional allocation (Martens & Wyble, 2010).

Early attempts to explain the AB hypothesised that it was due to resource limitations. These theories posited that the locus of the blink was a bottleneck at the stage of working memory consolidation, a stage subsequent to much perceptual processing (Chun & Potter, 1995). Consistent with this, behavioural and electrophysiological evidence shows that blinked stimuli are processed to a high level. Semantic aspects of blinked words makes the detection of subsequent, semantically-related, words more efficient (Shapiro, Driver, Ward, & Sorensen, 1997) and electrophysiological correlates of semantic processing are present for blinked stimuli (Luck, Vogel, & Shapiro, 1996; Vogel, Luck, & Shapiro, 1998), while those for working memory consolidation are attenuated for blinked stimuli (Vogel et al., 1998). However, findings that the blink is not present when participants report a sequence of several targets with no intervening distractors are difficult to explain in terms of resource limitations (Di Lollo, Kawahara, Ghorashi, & Enns, 2005; Olivers, van der Stigchel, & Hulleman, 2007).

Some recent theories explain the blink in terms of a suppressive mechanism. Exogenous attention is triggered upon detection of the target feature, sustained by the presentation of subsequent targets and suppressed upon processing of a distractor. This serves to protect working memory from encoding distractors (Olivers & Meeter, 2008) or to ensure the episodic nature of visual experience (Wyble et al., 2009).

SPEs have been measured in AB tasks to reveal that temporal selection during the blink is disrupted. Chun (1997) demonstrated that at lags that frequently yield the attentional blink, when the target was not reported the responses were often items presented near to the target in time. Popple and Levi (2007) explained such responses, and the AB generally, as errors in temporal binding. Vul, Kanwisher and Nieuwenstein (2008) assessed SPE distributions in response to blinked targets and observed that selection was delayed and more variable for targets presented during the blink. Goodbourn et al. (2016) analysed SPE data from multiple studies and fit mixture models to assess changes in temporal selection. These revealed that guessing increased during the blink and demonstrated a delay in selection that became smaller as the time between targets increased. Unlike the cruder analysis conducted by Vul, Kanwisher and Nieuwenstein (2008), mixture modeling did not find a change in the variance of selection.

The AB reflects changes in the ability to deploy attention over time and attentional selection during the AB is disrupted, but this does not tell us about the time taken to deploy attention in response to a cue. Early RSVP evidence about the latency of attentional deployment came from Lawrence (1971), who asked participants to report a word written in upper-case letters in an RSVP stream of lower-case words and found that they would often report the word following the target at presentation rates of 16-20Hz. Another relatively early investigation of attentional timing using RSVP is that of Reeves and Sperling (1986). In their study, participants viewed two simultaneous RSVP streams, one on either side (left and right) of a fixation point at eccentricities of .935º. The left stream consisted of letters and the right consisted of numerals. Participants were instructed to detect a target letter in the left stream and, upon doing this, report the earliest four numerals they could from the right stream while maintaining fixation. This method presumably requires a covert shift of endogenous attention from one stream to the other. The serial position of the numerals reported on each trial provides information about the timing of this attention shift.

The detection and report of several stimuli in quick succession in Reeves and Sperling's (1986) method is similar to the report of two targets in an AB design. However, given that participants can successfully report a sequence of targets if there are no intervening distractors, detecting a target and then shifting attention to a new set of targets may not result in the blink (Di Lollo et al., 2005). Consistent with this, Reeves and Sperling's (1986) participants often successfully reported items from the 200-500ms period during which the blink is typically observed. However there were errors in the order reported for the items, much like those observed in AB experiments where multiple successive targets must be reported and participants can report the identity of the targets, but not their temporal order (Chun & Potter, 1995; Wyble et al., 2009).

The participants of Reeves and Sperling (1986) reported numerals from 200 to 600ms after the onset of the target stimulus in the left stream. As mentioned, there were systematic errors in the serial order of items reported by participants. Participants tended to first report numerals that occurred around 400ms after the onset of the target, and subsequently reported items did not reflect the serial order of the numerals in the RSVP stream.

Reeves and Sperling (1986) explained their results with a theory of attentional selection in which attention operates like a gate. Attention is initially focused on detecting the target letter in the other RSVP stream. Upon detection of the target, attention is directed to the numeral stream. Attention gates access to central cognitive processes and operates like a gate. The attention gating function rises and falls over time, leading to the encoding of numerals from the stream. The strength attention at a particular time determines the strength of the representation of a numeral present at this time. The order of reported items is determined by the rank order of encoded numerals' activations. The first reported item is that which receives the most activation because its representation coincided with the time that attention was most active. Temporal errors arise because there is a variance in the strength of item activations. Items with low activation thanks to the low level of the attentional gate at the time they were processed have a poor signal-to-noise ratio and are likely to be reported in the wrong order.

In a similar investigation of the time course of attentional selection of multiple stimuli from an RSVP stream, Weichselgartner and Sperling (1987) used a method that did not require shifting attention between streams. Participants in that study viewed a single RSVP stream of numerals, presented at a rate of 10 or 12 Hz. One of these numerals was cued with an outline of a square or by virtue of having higher contrast than the rest of the stream and participants were instructed to report the target stimulus along with three subsequent stimuli. The probability of recalling an item from a particular time after the target had a bimodal distribution. The probability of recalling the target and the subsequent item (0 to 200 ms) was relatively high, but there was a reduction in reporting at 200 ms. From this point, a second peak emerged around 300-500ms after the onset of the target. To investigate the source of this bimodality, Weichselgartner and Sperling asked participants whether their reports seemed perceptually bound to the target. The early selections appeared perceptually attached to the target, whereas those from 300-500ms did not. The second, later, component had a similar temporal distribution as reports by the same participants generated from Reeves and Sperling's (1986) method.

Weichselgartner and Sperling (1987) argued that the distribution of responses in their single stream task was made up of two components. The first was an automatic process triggered by the detection of the target. This process was rapid. Participants reported the first two items with high accuracy, indicating that selection occurred over a period of less than 200ms from the onset of the target. The automatic process' timing is consistent with the peak timing of exogenous attention, which is often triggered with abrupt luminance increments like the cues that Weichselgartner and Sperling (1987) used (i.e. Jonides & Yantis, 1988). The second component reflected a voluntary process. It shared temporal properties with the voluntary deployment of attention prompted by Reeves and Sperling's (1986) task, and participants did not perceive these stimuli as perceptually attached to the target.

1.5 Temporal Selection in Vision: Buffering

These results demonstrate selection of stimuli presented after or with a cue in an RSVP stream, which is to be expected if the selection process is attention triggered by the cue. However, some results indicate that participants may report items from RSVP streams with a timing that is inconsistent with cued attentional selection. Specifically, it appears as if participants are making reports of items from before the time of the cue that are not guesses.

The inspiration for the current thesis came from Goodbourn and Holcombe (2015). Those authors were interested in the selection of simultaneous targets from multiple RSVP streams. They presented participants with two streams of letters, cued one or both of them with a white ring and calculated the SPE of participants' responses for each stream and trial. They observed robust effects of the spatial arrangement of streams on guessing rate when participants had to select two simultaneous targets. However it is their temporal results that are most relevant here. When selections were not a guess, the distributions of SPEs participants produced appeared to include items from before the time of the target. This pattern of results was replicated in Holcombe, Nguyen and Goodbourn (2017).

To explain the presence of pre-cue items, Goodbourn and Holcombe (2015) hypothesised that information about unattended stimuli is buffered so that it persists beyond the presentation of a subsequent stimulus. RSVP stimuli from different points in time are represented in this buffer and one is bound with the cue, a process that is distinct from attentional sampling from stream that is triggered by the cue. The bound representation is tokenised, consolidated into working memory, and reported. Selection from the buffer is error-prone and may sometimes result in a report of an item that was presented before the onset of the cue. Such responses are unlikely under an attention shift and thus constitute critical evidence for buffering.

1.6 Buffering is a problem for our understanding of visual selection

Goodbourn and Holcombe's (2015) data implied a buffer in which stimulus representations persist beyond the presentation of a subsequent stimulus, despite masking. This is at odds with assumptions regarding the effect of masking in models in which representations of unattended stimuli are terminated by masking (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002) or are briefly sustained, but are not available for selection (Olivers & Meeter, 2008; Wyble et al., 2009). The buffer, however, implies that unattended stimulus representations persist beyond the onset of a subsequent stimulus and that this information is available for selection. Evidence for pre-cue selections would require us to rethink the role of masking in RSVP by allowing information to be selected from sensory memory.

The presence of pre-cue items would also require us to rethink the way in which attention selects an item from an RSVP stream. It is typically assumed that stimuli from an

RSVP stream are not processed until the cue is presented and detected. The cue triggers attention, which samples an item from the stream (Chun & Potter, 1995; Olivers & Meeter, 2008; Weichselgartner & Sperling, 1987; Wyble et al., 2009). However, buffering suggests that there is processing of stimuli prior to the onset of the cue because without this, pre-cue stimuli could not be reported.

The visual world is dynamic and full of stimulation. Attention allows us to flexibly allocate our visual resources to stimuli within a scene, but what we observe at one point may be masked by stimuli at the same visual location when we make a saccade, move, or there is movement in the scene. RSVP, with its inherent masking, mimics this process of attending to relevant stimuli in a stream of visual information, but in a controlled environment. Theories about how this is achieved are violated by buffering, which suggests that stimuli are processed regardless of their relevance and that sensory information is resilient in the face of masking. Investigating how attention can select relevant information from buffering despite masking provides key information regarding how attention samples the visual world.

1.6.1 Assessing buffering.

Goodbourn and Holcombe (2015) were not the first to observe reports of stimuli from before the time of the cue in RSVP, and to suggest that these reports occur more often than chance. Such observations also appear in previous literature on the time course of selection in RSVP both with a single cue (Botella, 1992; Botella & Eriksen, 1991; Gathercole & Broadbent, 1984; McLean, Broadbent, & Broadbent, 1983) and in the attentional blink literature (Vul, Nieuwenstein, et al., 2008). However, this thesis describes the first quantitative test of such responses. The published evidence for pre-cue selections, which will be described in more detail in Chapter Four, relies on observations that there appear to be more reports of stimuli from before the cue than could be expected by guessing alone. However none of the cited papers provide a test of whether or guessing may explain these reports. To distinguish between guesses and non-guesses one must first estimate the rate of guessing in the task, and secondly develop some statistical measure of the deviation between the proportion of responses from before the cue and the same proportion as predicted by guessing.

The first issue is what we refer to as estimating the *efficacy* of selection given the cue. That is, how well participants can detect the cue and accurately identify a letter from around that time. In each trial, the RSVP items are presented in a random order. Failing to detect the cue or misidentifying a letter will lead to reports of stimuli that are uniformly distributed in time relative to the cue because of this random presentation. This can lead to reports of items from before the time of the cue, and the presence of responses of items from before the cue is critical evidence for buffering. We thus need some way to account for non-efficacious responses.

What is needed is to estimate the proportion of trials in which participants made an efficacious response - that is, where they identified a letter from around the time of the cue. Mixture modelling achieves this (Zhang & Luck, 2008). In mixture modelling applied to SPEs, the distribution of SPEs is modelled as a mixture of two probability distributions (Goodbourn & Holcombe, 2015). One of these distributions represents efficacious responses and the other represents non-efficacious responses - where the cue was missed or the selected letter misidentified. Mixture modelling of SPE distributions in this manner allows us to estimate the proportion of responses that were efficacious and those that were not. This also allows us to estimate the temporal properties of the efficacious distribution - the mean and standard deviation of selection in time after accounting for responses that were not efficacious. I will discuss mixture modelling more in Chapter 2.

The second necessary condition for detecting buffering is a statistical procedure that allows us to test whether responses from before the cue are more frequent than would be expected by guessing. Goodbourn and Holcombe (2015) observed that the distribution of SPEs expected for responses that are not efficacious is a windowed uniform distribution. This insight allowed us to develop a test for deviations from this distribution at SPEs before the time of cue, based on the observation that the frequency of responses at a particular SPE relative to those at all other SPEs will have a binomial distribution. Doing so allows us to generate a probability for the count of responses at a particular SPE if responses there are not efficacious.

1.7 Aims

This thesis is an examination of the temporal properties of visual selection using RSVP, mixture modelling, and the binomial test. The main aim is to investigate the presence of buffered information. When do we see responses of RSVP stimuli that were presented before the time of the cue? This is important because many models of visual selection from RSVP streams assume an exogenous attentional component, triggered by the cue and termination of unattended item representations by masking. These factors should not result in pre-cue reports that are efficacious, but the results observed by Goodbourn and Holcome (2015) and Holcombe, Nguyen and Goodbourn (2017) suggest that such reports exist. This implies that attention can select information about the rich stream of visual information we observe in naturalistic viewing from sensory memory, despite masking.

The thesis is also concerned with the temporal properties of visual selection. What is the latency and variability of selection across eccentricity, with central or peripheral cues and under conditions in which the cue may interfere with the target? Investigating these questions using mixture modelling and the binomial procedure allows us to account for trials in which the cue was missed or the selected letter was misidentified, and instead estimate the properties of successful visual selection from an RSVP stream. Doing so provides the first direct evidence of buffered responses in RSVP.

The second chapter of the thesis will describe mixture modelling fitting and the binomial test, as well as our attempts to validate the mixture models with parameter and model recovery. Chapter 3 investigates voluntary contributions to buffered reports with an experiment comparing selection distributions when the cue is a central, symbolic cue and those in which the cue is a peripheral, spatial cue - a white ring at the location of the target. Chapter 4 describes changes in the frequency of buffered items with changes in the number of simultaneous RSVP streams, and provides evidence that buffered reports are affected by the number of monitored streams. In Chapter 5 we test for changes in the temporal qualities of selection across eccentricity, because previous investigations of vision’s temporal properties indicate that visual processing may be faster outside the fovea.

Chapter 2: Analyses

All the experiments in this thesis use RSVP and all analyses are based on the serial position error (SPE) of the stimulus reported by participants. The SPE refers to the temporal position of the reported stimulus, in item units, relative to the target on a particular trial (Figure 1). The SPE gives a quantised measure of the timing of selection on a particular trial, so over many trials the distribution of SPEs produced by a participant under particular stimulus conditions (i.e. the number of streams, the kind of cue, or the eccentricity of the target) provides information about the temporal qualities of selection under those conditions.

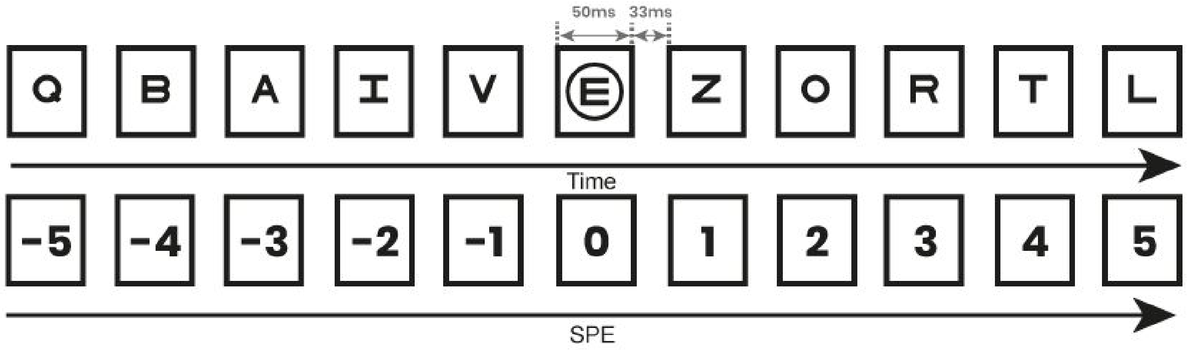

Figure 1. The top row shows a schematic example of an RSVP stream of letters, presented at one spatial location, with each stimulus presented with a stimulus onset asynchrony of 83 ms (roughly 12 Hz). The target stimulus is the letter “E”, as indicated by its co-occurrence with the cue, a ring around the letter. Each stimulus is associated with an SPE representing its serial position relative to the target stimulus.

However, the distribution of SPEs does not only include responses that were informed by the cue. Instead, it is a mixture of responses resulting from trials in which participants misidentified a selected letter or failed to detect the cue and responses from trials in which the participant selected a letter based on the timing of the cue. The former sorts of responses are errors, which we term *identification failures* because they result from a failure to identify a letter or the cue. These failures do not tell us anything about the temporal qualities of selection in RSVP, but we want to account for them in order to investigate the temporal characteristics of *efficacious selections*, those responses that are based on the timing of the cue.

This distinction between efficacious responses and identification errors is critical to the analyses we use in this thesis. In this chapter I describe two analyses based on SPEs, one of which is novel. The first, mixture modelling (Zhang & Luck, 2008), allows us to estimate the proportion of SPEs resulting from identification failures and efficacious selections. In doing so, the mixture modelling procedure allows us to make inferences about efficacious selection's temporal qualities, its distribution, mean and standard deviation. The second analysis we describe is a test for whether the number of responses with a certain SPE - that is, at a certain time - is more frequent than would be predicted based on identification failures alone. This allows us to test for the presence of buffered responses - efficacious reports of stimuli from before the cue. To do this we use the distribution of SPEs expected under identification failures as the null hypothesis for a test statistic.

2.1 Mixture Modelling and Model Comparison

We model the distribution of SPEs as the mixture of two distributions using the mixture modelling procedure described by Goodbourn and Holcombe (2015). The observed distribution of SPEs generated by a participant is assumed to be the output of two processes: efficacious responses, or responses informed by the cue, and identification failures. The latter component represents reports that were guesses because the participant did not detect the cue on that trial, or errors caused by participants misidentifying the selected letter. We fit two models to each participant's data, one representing buffering and the other representing attention shifts. Each model consists of a mixture of the distribution of identification errors with a distribution that reflects efficacious responses (Zhang & Luck, 2008). We fit the models using maximum likelihood estimation in a custom R package (https://doi.org/10.5281/zenodo.3545085) and compare their fits in order to assess the relative evidence for buffering and attentional selection.

Both models, the model representing buffering and the model representing attention shifts, contain a component representing identification failures. When participants make a response that is the result of an identification failure, their response is unrelated to the time of the cue and thus can come from any SPE with a uniform probability. As explained by Goodbourn & Holcombe (2015), the fact that the serial position of the cue differs across trials means that the minimum and maximum possible SPEs varies across trials as well. For instance, when the cue is in the sixth of 24 of serial positions, the maximum SPE is 18, but when it is in the 10 serial position the maximum SPE is 14. This means that the distribution of SPEs is tapered at its extremes, due to changes in the serial position of the cue. In practice, this windowing mainly affects the distribution of identification errors, because those may result in very extreme SPEs, whereas efficacious responses appear to be concentrated within a few hundred milliseconds of the cue. Thus the distribution of identification failures is a windowed uniform distribution (Figure 2).

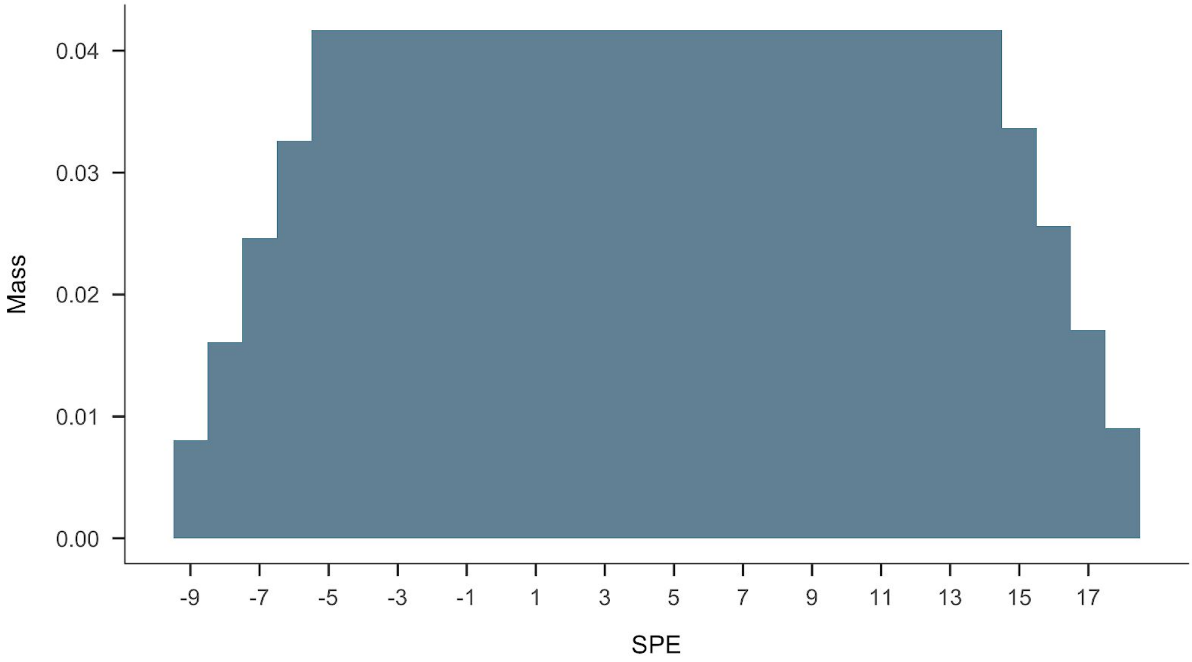

Figure 2. The distribution of responses associated with identification errors.

We assume that responses produced by attentional selection or buffering impart distributions with different shapes to the observed temporal distribution. Our mixture modelling procedure fits models with different shaped efficacious distributions in order to detect these shapes in the observed SPE distributions (Figure 3).

The shape of the efficacious distribution should differ depending on whether items are selected with an attention shift or from the buffer (Figure 3). An attention shift should produce a positively skewed distribution with no responses of items from before the cue. This is because the onset of the cue triggers the shift, imposing a lower bound on selection times so that variance in selection times will be distributed in the right tail of the distribution. The SPE distributions that have been interpreted as evidence for buffering, on the other hand, appear symmetric (Goodbourn & Holcombe, 2015; Holcombe et al., 2017).

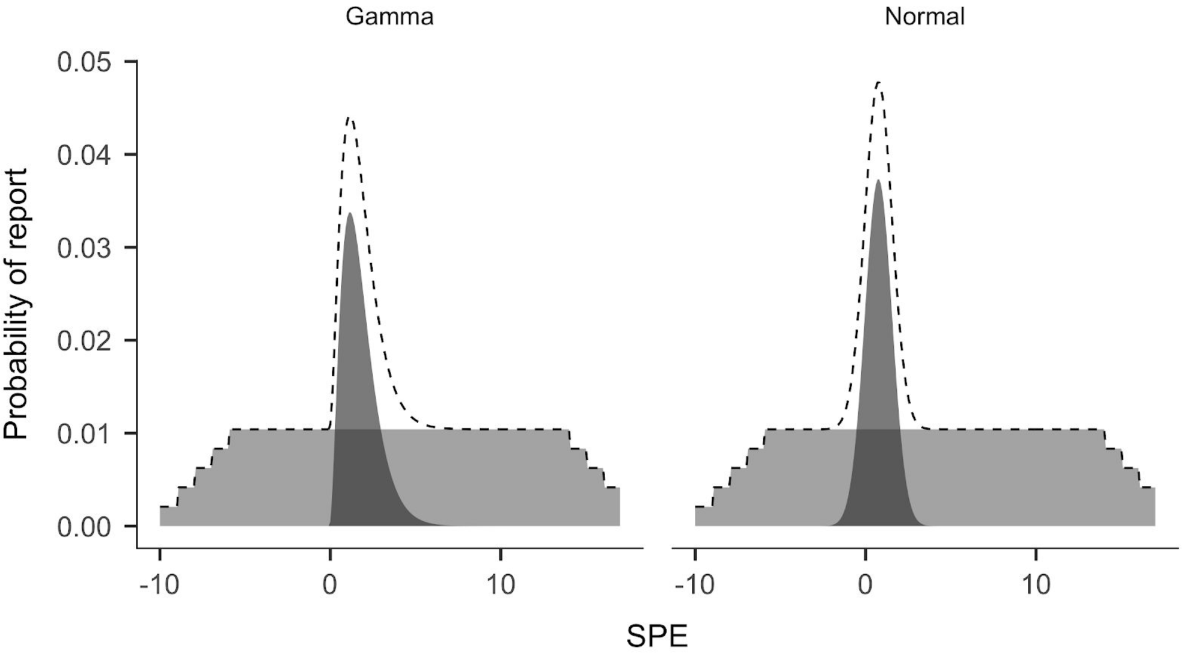

Figure 3. Example mixtures for the different models. Both models combine a distribution of identification failures (the light grey shaded area) with a distribution of efficacious selections (the dark grey shaded area). The dashed line represents the probability of report resulting from summing these distributions. The efficacious distribution differs between the models. In the left model, the efficacious distribution is a gamma distribution, which we use to model attention shifts. The right model represents buffering by using a Gaussian distribution as the efficacious distribution.

We model attentional selection using a gamma distribution[1], which has positive skew and does not produce responses from before the cue. For the buffering model, the assumed efficacious distribution is a Gaussian, as in Goodbourn and Holcombe (2015). Combining these efficacious distributions with the distribution corresponding to identification failures produces two mixture models, a symmetric model where the efficacious distribution is a Gaussian distribution (Figure 3, right side), and a skewed model where the efficacious distribution is a gamma distribution (Figure 3, left side).

The Gamma distribution has two parameters: shape (α) and scale (θ) (Thom, 1958). Its probability density function is given by,

$$Gamma(x, \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\theta}}$$

Where Γ(α) is the gamma function. This density function is defined for non-negative values of x only and has a positive skew controlled by the shape parameter (α), where smaller values of the shape parameter lead to more skew. The mean of the distribution is,

$$M = \alpha\theta$$

Its standard deviation is,

$$SD = \sqrt{\alpha}\theta$$

---

[1] We also attempted to model this distribution with a log-normal and a Gaussian truncated to provide skew. The log-normal mixture models performed poorly in a parameter recovery test, and the Gamma was preferred to the truncated Gaussian because the latter required an arbitrary choice about truncation point's location

Combining the gamma distribution with the uniform creates a mixture model $h()$ with three parameters: α, θ and p,

$$h(x; p, \alpha, \theta) \; = \; W(x)[(\tfrac{p}{C_g})Gamma(x, \alpha, \theta) + (\tfrac{1-p}{C_u})U(x)]$$

Here, x is the time of the response relative to the cue in SPEs, W(x) is the windowing function, described in Goodbourn and Holcombe (2015), which tapers the mixture model near its extremes to account for the fact that extreme SPEs are not possible on every trial due to changes in the temporal position of the cue. The remaining elements of the model are: α and θ, the parameters of the gamma distribution; p, the probability of an efficacious response, and U(x), the uniform distribution. The values C_u and C_g are normalising constants for the uniform and gamma distributions, respectively, so that the integral of the distribution is equal to one.

The Gaussian mixture model, described in detail in Goodbourn and Holcombe (2015), combines the uniform distribution with a Gaussian. This allows efficacious responses from before the cue, because unlike the gamma distribution the Gaussian is defined for negative numbers. It is also symmetric, a quality associated that appears to be associated with buffering, although this has yet to be tested (Goodbourn & Holcombe, 2015; Holcombe et al., 2017). The model has the form,

$$f(x; p, \mu, \sigma) \; = \; W(x)[(\tfrac{p}{C_n})N(x, \mu, \sigma) + (\tfrac{1-p}{C_u})U(x)]$$

Where N(x,μ, σ) is the normal distribution, μ and σ are the mean and standard deviation of the normal, and C_n is the normalising constant for the normal distribution. All other elements are the same as those in the gamma-uniform mixture.

We fit both these models using maximum likelihood estimation. For each kind of model, this process yield three parameter estimates: *Efficacy*, the proportion of reports that are efficacious, *Latency,* the mean time of the efficacious distribution relative to the cue, and *Precision*, the standard deviation of the efficacious distribution. Latency and precision are thus estimates of the temporal properties of selection, whereas efficacy estimates the extent to which participants can detect the onset of cue and identify a letter from around that time.

To compare how well the two different models fit each participant’s data, we used the Bayes factor – the probability of the data under one model divided by the probability under the other model. The Bayes factor, with an uninformative prior over the model parameters, can be estimated using the Bayesian information criterion for each model (Raftery, 1999; Wagenmakers, 2007). We calculate the BIC for each model, and the bayes factor estimated from the BICs to assess which model is a better fit to each participant’s data. This mixture model comparison gives us a ratio for the evidence of one model relative to the other. Bayes factors with a ratio greater than 3 in favour of a particular model are taken as evidence for that model.

2.2 Model and Parameter Recovery

To test how well our mixture model could estimate the parameters of an SPE distribution, we simulated participants with different efficacies, latencies, precisions, and efficacious distributions (gamma or Gaussian) and fit the mixture model to these simulated response distributions. Generating data where the ground truth was known in this manner allowed us to test the error in parameter estimates associated with the maximum likelihood estimation procedure. It also allowed us to assess the ability of the model comparison procedure to discriminate between these two shapes of the efficacious distribution.

The parameters with which we simulated data are presented in Table 1. Latency and precision values were chosen based on the range of parameter estimates observed in Goodbourn and Holcombe (2015). Latency parameters were chosen based on the latency one may expect for an exogenous attention shift (120 ms = .84 SPE with an 83ms SOA; Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989) with a longer value included to test a range of latencies. Each parameter value was combined with the values of all other parameters in the table. For each combination of efficacy, latency, precision and efficacious distribution, we simulated 100 participants, resulting in 1600 simulated participants, with 200 trials per participant.

Table 1

Parameter values for simulations. Latency and precision are in SPEs. All combinations of parameters and efficacious distribution were tested

Efficacy	Latency	Precision	Efficacious distribution
.7	1.5	.84	Normal
.9	2	1.2	Gamma

The first step for simulating a trial involved sampling a cue position in time over the range we use in our task (serial positions from 6 to 10, inclusive). Next, we determined whether or not the simulated trial was efficacious (probability = efficacy). If the trial was efficacious, we sampled an SPE from the efficacious distribution being used by that simulation (gamma or Gaussian). If the sampled SPE was outside the possible range of SPEs in our task, we sampled the SPE again. For the Gaussian efficacious distribution, latency and precision were the mean and standard deviation, respectively. For the gamma distribution, latency and precision needed to be transformed into shape and scale, that distribution's parameters. The shape and scale of the gamma distribution were calculated by the equations below, which follow from those provided for the mean and standard deviation of the distribution:

$$Shape = \frac{Latency^2}{Precision^2}$$

$$Scale = \frac{Precision^2}{Latency}$$

If the trial was not efficacious and thus was an identification error, we sampled an SPE from a uniform distribution with bounds at the minimum and maximum possible SPE on that trial, given the temporal position of the cue.

Once we had simulated a set of 200 trials, we fit the mixture models to the simulated data. For each simulated participant, which had efficacious trials drawn from only one kind of efficacious distribution, we fit both models.

2.2.1 Results

Model recovery. We calculated Bayes factors to assess which model best fit a set of simulated data. A Bayes factor that indicated that the data favoured a particular model over the other by a ratio of three or greater was taken as evidence for that model. There are three outcomes for a particular pair of models fit to the same data. If the Bayes factor fell between .33 and 3, we concluded that the evidence for the models was equivocal and we could not decide in favour of either model (Jeffreys, 1998). If the evidence is no equivocal, the Bayes factor may favour the correct model, which we term the “generative” model, because it was the kind of distribution that generated the data. The Bayes factor may otherwise favour the wrong model (the distribution that did not generate the data). The proportion of model fits that fall into either of these categories is presented in Table 2. These proportions do not consider differences in efficacy, latency or precision. The model comparison procedure identified the correct model in over 90% of cases and was more likely to be unable to facilitate a decision than to facilitate the wrong decision.

Table 2

Model comparison results.

Generative Model	Correct Model	Incorrect Model	Neither Model

Gamma	.915	.023	.062
Gaussian	.918	.008	.075

Parameter estimates.

We report the parameter estimates based on the model that generated the data and the model fit to the data. This allows us to assess not only how well the correct model can recover the parameters associated with the data, but any deviations in parameter estimation that can be expected by analysing data based on a model fit that does not match the generative model. Fortunately, such deviations are small, even when the generative model and the model fit to the data do not match.

The deviation between the parameter estimates produced by the model fitting and the parameters used to generate the data is calculated as

$$Error = Parameter\ Estimate - Generating\ Parameter$$

Negative errors mean that the model fitting procedure underestimated the value of a parameter. Positive errors mean that the value was overestimated.

Efficacy. Efficacy estimates showed only small errors. Efficacy errors are summarised in Table 3 for the different generative models and the models fit to the data. The data are also presented in Figure 4. These values do not account for differences in the values of latency and precision used to generate the data. The error in efficacy estimates was indistinguishable from zero when the generative model and the analysis model matched. When there was a mismatch between the models, efficacy was underestimated, however this was only a small underestimate.

Table 3

Summary statistics for efficacy estimate errors and quantiles

Generative Model	Model fit to data	Mean Error	SD	2.5% Quantile	97.5% Quantile
Gamma	Gamma	0.001	0.031	-0.057	0.063
Gamma	Gaussian	-0.017	0.033	-0.083	0.052
Gaussian	Gamma	-0.014	0.038	-0.083	0.057
Gaussian	Gaussian	0.003	0.034	-0.063	0.061

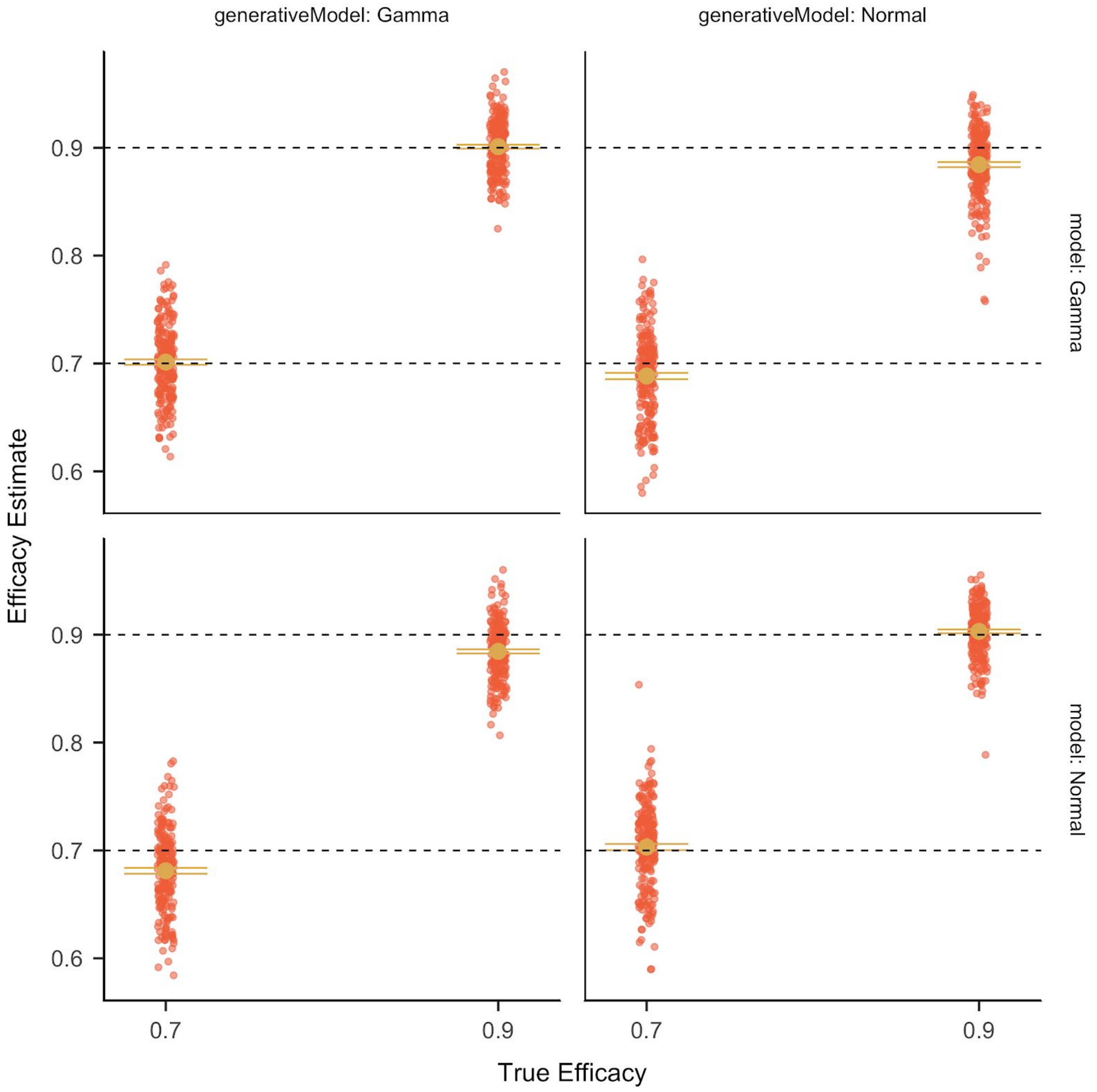

Figure 4. Efficacy estimates based on the true efficacy used to simulate the data, the generative model, and the model fit to the data. The columns of the grid represent the model used to generate the data. The rows represent the model fit to the data. The orange points are the estimates for each

simulated participant, with a small amount of horizontal jitter. The yellow points and error bars represent the mean±SE. The dashed lines extending from the y axis show the true efficacy values for reference.

Latency. Latency error was also small. The error for this parameter is summarised in Table 4 and Figure 5, where it is presented in SPEs. When the generative model and the model fit to the data match, there is very little error. In this case the true value and the estimate deviate by less than .01 of an SPE. When there is a mismatch between the generative model and the model fit to the data, the deviation between the estimate and the true value is about .1 of an SPE. This corresponds to less than ten milliseconds error with the stimulus onset asynchronies we use in our experiments (66 to 83 ms).

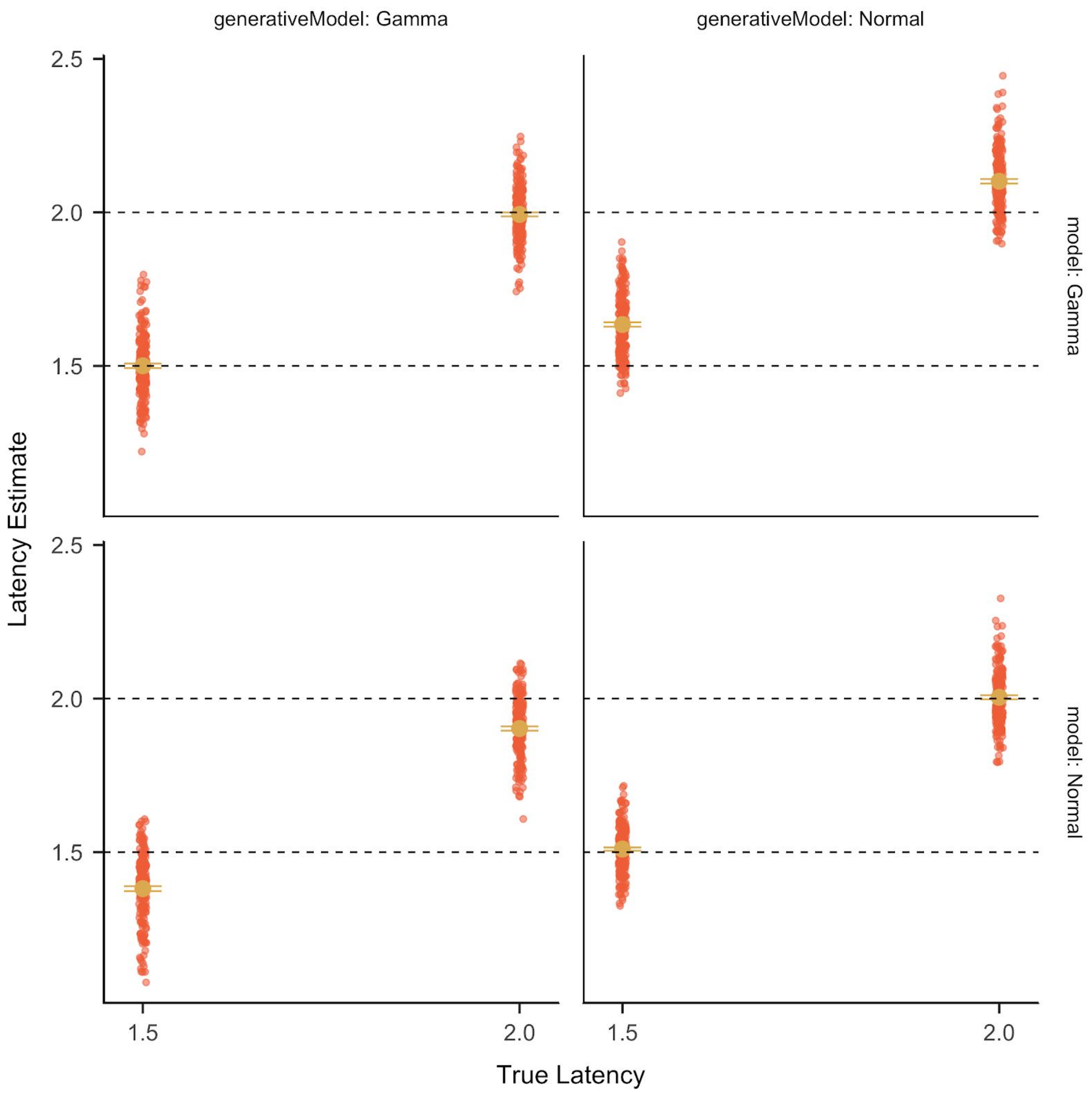

Figure 5. Latency estimates, in SPE, based on the true latency used to simulate the data, the generative model, and the model fit to the data.

Table 4

Summary statistics for latency estimate errors in SPEs

Generative Model	Model fit to data	Mean Error	SD	2.5% Quantile	97.5% Quantile
Gamma	Gamma	-0.003	0.098	-0.183	0.212
Gamma	Normal	-0.108	0.107	-0.343	0.089
Normal	Gamma	0.118	0.104	-0.06	0.339

Normal	Normal	0.007	0.085	-0.156	0.172

Precision. A similar pattern of errors was found for precision (Table 5, Figure 6). Error was small, less than .01 of an SPE, when the generative model and the model fit to the data matched. When the generative model and the model fit to the data did not match, the error was larger, but even the largest mean error (-0.13) does not correspond to more than 11ms with the presentation rates used in the experiments described in this thesis.

Table 5

Summary statistics for precision estimate errors in SPEs

Generative Model	Model fit to data	Mean Error	SD	2.5% Quantile	97.5% Quantile
Gamma	Gamma	-0.006	0.109	-0.203	0.232
Gamma	Normal	-0.132	0.111	-0.368	0.051
Normal	Gamma	0.058	0.105	-0.119	0.274
Normal	Normal	-0.008	0.08	-0.165	0.156

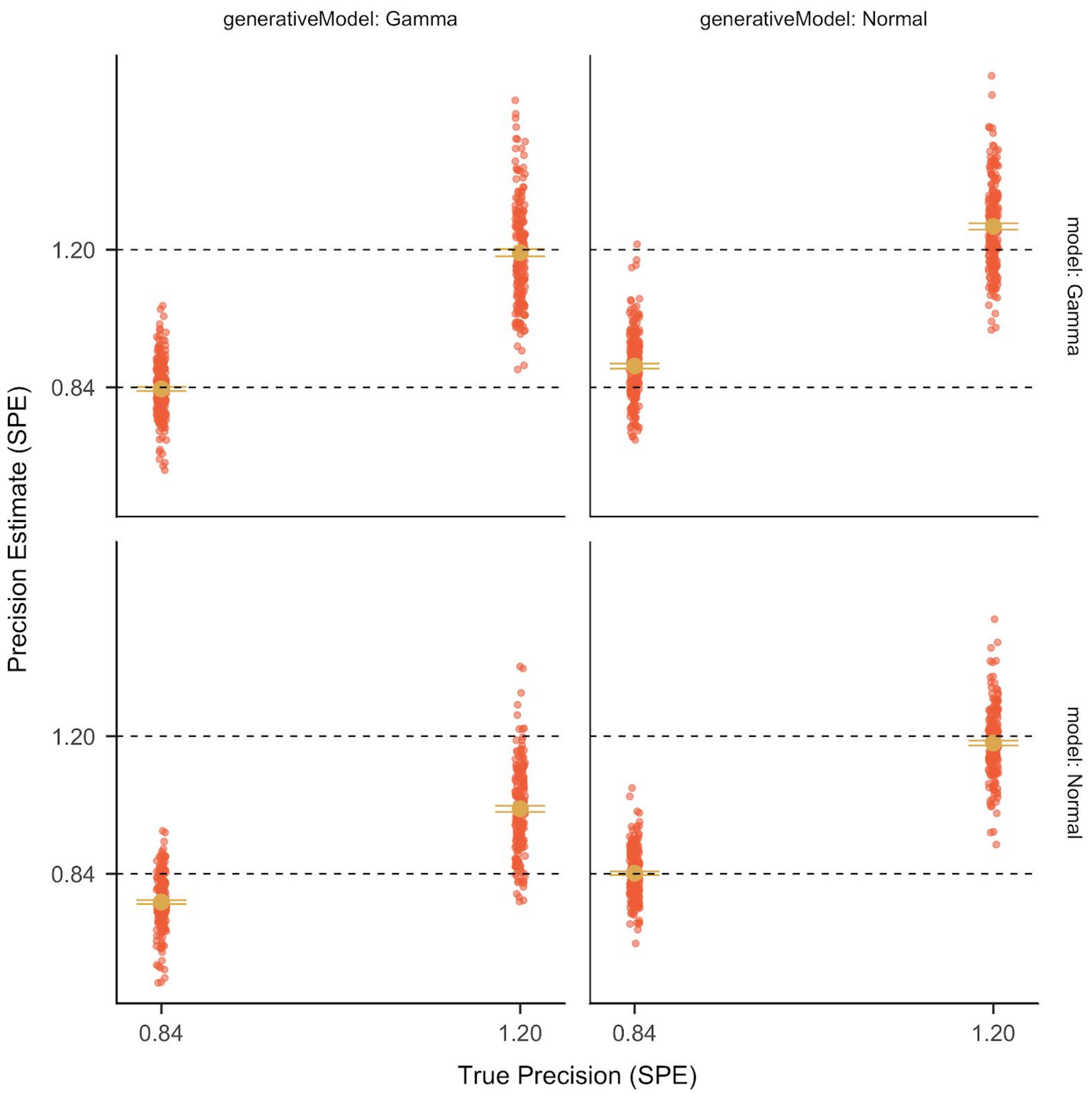

Figure 6. Precision estimates, in SPE, based on the true latency used to simulate the data, the generative model and the model fit to the data.

2.2.2. Discussion.

This parameter recovery study indicates that the maximum likelihood estimation procedure that we use to fit mixture models to our data and our model comparison analysis performs well. The model comparison identifies the true efficacious distribution in approximately 90% of cases. Parameter estimation results in errors with small magnitudes.

Errors are largest when the generative model and the model fit to the data do not match, but the errors resulting from such mismatches are small. In the temporal domain they correspond to 1 to ten ms with the presentation rates used in our designs, and for efficacy they result in an underestimate that is at most approximately .02.

2.3 The Binomial Test

Previous studies provided no direct evidence for efficacious pre-cue selections. They relied on the indirect evidence of a best-fitting Gaussian mixture model having latency and precision estimates such that the Gaussian included some responses that were before the cue (Goodbourn & Holcombe, 2015). We developed a more direct test, which we call the binomial test, for efficacious reports of stimuli before the cue.

The binomial test provides a liberal estimate of the probability of a particular number of pre-cue reports assuming that such reports are due to identification failures. This provides a conservative null hypothesis test for each SPE – if the data reject it, we conclude that efficacious responding contributed to that serial position. That is, we use the binomial test to assess whether the observed count of SPEs at -1 (the item presented before the target) is hard to account for ($p<.05$) with identification failures. The SPE = -1 response is a particular proportion of the pseudo-uniform identification failure distribution. Under the null hypothesis of all responses being identification failures, this proportion is the probability of making an SPE=-1 response, and the binomial distribution indicates the probability of any particular count of SPE=-1 responses, given the total number of trials. This allows calculation of a p-value that tells us the probability of observing a count of SPEs at -1 at least as extreme as the observed count if participants' responses are entirely comprised of identification failures. We will illustrate this before making the test more realistic by not assuming that all responses are identification failures.

To illustrate the binomial test, the distribution of selection errors of a participant from Chapter 4's first experiment is shown in Figure 7A. Identification failure responses have a nearly uniform distribution of SPEs, and SPEs of -1 make up some proportion of this distribution (Figure 7B). To test whether there are more responses at -1 than expected from identification failures, we use the binomial distribution, where hits are SPEs of -1 and misses are all other SPEs (Figure 7C). The proportion of the identification failure distribution at -1 is used as the probability of a hit in this distribution and the number of responses recorded by a participant as the number of observations. The p-value is the orange region of Figure 7C – the probability of a count of -1 SPEs at least as extreme as the observed count in the binomial distribution. As discussed below, we adjust the identification distribution to account for the fact that participants' SPE distributions are unlikely to be entirely due to identification failures. We interpret a p-value below .05 as good evidence against the null hypothesis that the SPE=-1 responses are entirely due to identification failures.

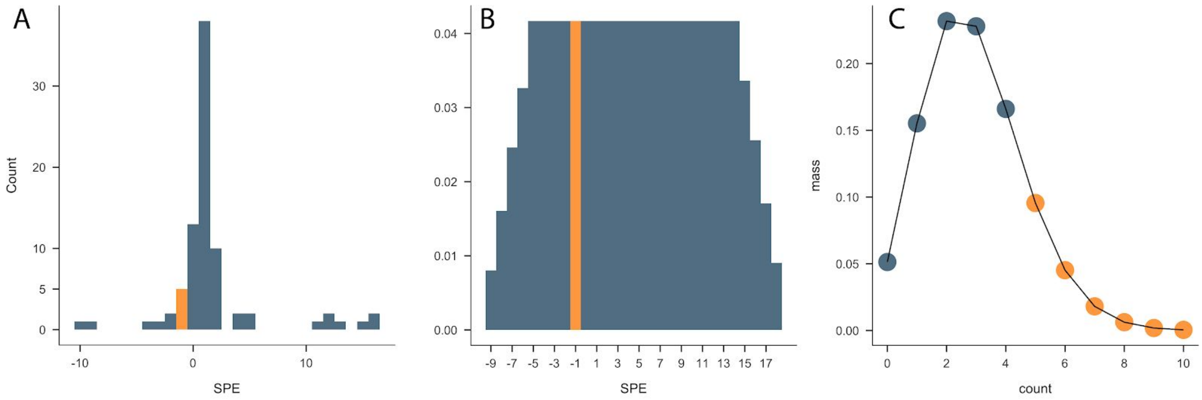

Figure 7. The process for testing whether the count at a particular SPE is greater than expected by identification failures. Panel **A** shows one participant's SPE histogram. The highlighted bar shows the five SPEs of -1 in this histogram. Panel **B** shows the distribution of SPEs associated with identification failures; the highlighted component is the proportion of the distribution corresponding to an SPE of -1. The points and lines of panel **C** show the binomial distribution associated with an SPE of -1 using the proportion of the identification failure distribution at -1 as the probability of success. The highlighted points in panel **C** show the proportion of the distribution with a count at least as extreme as that

shown in panel **A**. The probability of a random draw in this highlighted region corresponds to the p-value for the highlighted component in panel **A**, according to the test.

2.3.1 Making the binomial test less conservative

The distribution of selections is extremely unlikely to be comprised entirely of identification failures because previous work has repeatedly found that selection error distributions depart from the identification failure distribution. They have a peak around the time of the target, indicating that in some trials participants successfully identify the target and suggesting that they sometimes identify a letter near the time of the target (Goodbourn & Holcombe, 2015; Holcombe et al., 2017; Vul, Nieuwenstein, & Kanwisher, 2008; Martini, 2013). Using their Gaussian mixture model, Goodbourn and Holcombe (2015) estimated that identification failures (which they referred to as “guessing”, although the concept also includes misidentifications) accounted for about a quarter of their participants’ responses.

Given that not all responses are identification failures, the assumption of the binomial test (as described so far) that all responses are identification failures is wrong, and this makes the test overly conservative. That is, by assuming that all responses are identification failures, the probability we assign to a particular count of SPE = -1 responses is too high because the identification failure rate assumed, 1, is higher than the actual value.

To make the test less conservative, a more realistic estimate of the proportion of responses that are identification failures is needed. However, we should avoid using an estimate expected to be lower than the actual value, because that would result in a liberal test (one for which the p-values are too low). To account for the fact that not all responses are identification failures we need some proxy for efficacy. This proxy for efficacy should be, if it is wrong, an underestimate to avoid yielding an identification failure rate that is too low, which would result in a liberal test.

We used the proportion of responses with SPEs of 0 as a conservative estimate of efficacy and its complement, 1 - p(SPE = 0), as a liberal estimate of the rate of identification failures. The identification failure rate liberal estimate is used to make the estimated proportion of SPEs of -1 expected from identification failures more realistic. This more realistic estimate of the proportion of trials with an SPE of -1 for a particular participant is the product of the proportion of the identification failure distribution at -1 for that participant and the estimate of the failure rate. This adjusts the proportion of identification failures down from an unrealistic 100%. Doing so does not underestimate the proportion of non-efficacious responses, because efficacy estimated in this manner is conservative, as shown below. In this way we do not make the likely-untrue assumption that all responses are guesses.

How can we be sure the proportion of responses at zero is a conservative estimate of efficacy? One can imagine conditions in which this would not be true. Firstly, it could happen that all efficacious responses might fall at zero. In this case the observed count of responses at this point over-estimates efficacy, because the observed count includes all efficacious responses plus some of the non-efficacious responses. Another undesirable situation is that in which all efficacious responses fall somewhere other than zero, in which case the count at zero contains only identification failures, and our measure of accuracy actually represents the proportion of guesses at 0. These two scenarios represent extreme conditions under which we cannot use the count of responses with an SPE of 0 as a measure of efficacy. In other words, we want efficacious responses to be spread over multiple SPEs, one of which is zero.

Finally, the combination of efficacious responses and identification failure responses at zero may overestimate efficacy even if neither of the extreme conditions are met. In this situation the proportion of efficacious responses at zero and the proportion of identification failure responses at that SPE sum to greater than efficacy, in which case efficacy is overestimated and guessing is underestimated by this measure.

To investigate whether or not these undesirable conditions are present in an RSVP experiment similar to our own, we used data from the two-stream single-target condition of Goodbourn and Holcombe (2015). The experiments described in the current thesis involve multiple simultaneous RSVP streams, one of which is cued at one point in time. In Goodbourn and Holcombe's two-stream single-target condition, participants saw two RSVP streams, one of which was cued with a white ring indicated the target letter. The analyses reported in that paper for this condition analysed responses based on the spatial position of the cued stream in order to investigate potential spatial asymmetries in selection. However, there were no effects of spatial position on the efficacy, latency or precision of selection. We therefore collapsed the streams together and analysed the distribution of selections in the cued stream, regardless of the spatial position of that stream on particular trials. This is the same analysis we use in the present experiments.

The proportion of efficacious and identification failure responses at zero can be estimated by fitting a mixture model to the SPE data. Doing so allows us to investigate whether SPE = 0 is a conservative measure of efficacy in a dataset that is independent of the data used in the present experiments. After fitting mixture models to the Goodbourn and Holcombe data, we looked at the proportions of the efficacious and identification components of the mixture corresponding to an SPE of 0. The proportion of the efficacious distribution at SPE = 0 for each participant was, on average, 0.44 (SD = .1, min = .2, max = .7). The proportion of the total number of identification failures at 0 was 0.042 for each participant. The proportion at zero is the same for each participant because they each observed the same distribution of cue times (Goodbourn and Holcombe, 2015). Summed, the proportions of estimated efficacious responses and estimated identification failures at an SPE of zero did not exceed efficacy for any participant in this experiment.

The majority of responses at zero are, according to this analysis, efficacious and any non-efficacious responses represent only a proportion of the total non-efficacious distribution for each participant. Summing the proportion of efficacious and non-efficacious responses at an SPE of zero does not overestimate efficacy for any of the participants in this analysis. The conditions necessary for using the proportion of SPEs of zero as a conservative estimate of efficacy are met.

The complement of this conservative estimate of efficacy, $1 - p(SPE = 0)$, is a liberal estimate of non-efficacious responding. That is, while this allows us to account for the fact that not all responses are likely to be non-efficacious, it does not underestimate the rate of non-efficacious responding.

Given a particular participant's efficacy (the proportion of trials where $SPE = 0$), we assume that the probability of a non-efficacious response with an SPE of -1 was $p(SPE=-1|\text{non-efficacious}) * (1 - p(SPE = 0))$. For each participant, we generated a binomial distribution predicting SPEs at - 1 with this proportion. From this distribution we calculated a p-value for the observed count of -1 SPEs. The p-value is the proportion of the binomial distribution with a count equal to or greater than the observed count, given that the probability of an SPE at -1 is that expected by identification failures.

2.3.2 Model comparison and binomial analysis of data from Goodbourn and Holcombe (2015)

As discussed in the introduction, Goodbourn and Holcombe (2015) modelled their participants' SPE distributions (generated in a task with two simultaneous RSVP streams) with a Gaussian-uniform mixture. The mixture model parameter estimates produced by maximum likelihood estimation resulted in the appearance that the efficacious distribution contained a substantial proportion of items presented before the cue. However, it is not clear if the Gaussian mixture best fits the data. It may be that the efficacious distribution is not characterised by symmetry and some proportion of pre-cue items, but instead is skewed and post-cue. We hypothesise that attentional selection will result in a distribution with these qualities, represented by our gamma-uniform mixture model. To investigate whether buffering (the Gaussian mixture) or attentional selection (the gamma mixture) best explain these data, we fit the mixture models to the single target condition of Goodbourn and Holcombe's (2015) Experiment 1.

In this condition, participants saw two RSVP streams at a rate of 12 items/second. On each trial, one stream was cued with a white ring and participants were instructed to report this letter. We chose this condition over the others because the other conditions in this experiment involved simultaneous selections when two cues were presented and thus attention was divided over two locations, but in this condition attention was cued to a single location. The temporal characteristics of selection were found to be the same in the dual-target condition. However there is an efficacy deficit in the rightmost or inferior stream with two targets, depending on the spatial arrangement (Goodbourn & Holcombe, 2015). To analyse a condition in which the probability of an efficacious selection is high, which provides more data for the efficacious distribution fitting, we chose the condition without divided attention.

There were 26 participants (5 psychophysically-experienced) in this condition. Because Goodbourn and Holcombe (2015) found no spatial effects in this condition, we ignored the spatial location of the cue and collapsed the SPE data from the left or right stream into a single stream. We fit the buffering and attention mixture models and analysed the frequency of responses with SPEs of -1 with the binomial test.

The SPE distributions of twenty-three participants were best fit by the Gaussian mixture model, indicating buffering, one participant's SPE distributions was best fit by the skewed gamma mixture model, and two participants' SPE distributions were favoured by neither model. Thus, the majority of participants in this experiment, who produced data interpreted by Goodbourn and Holcombe (2015) as evidence for a buffer, had SPE distributions that were symmetric rather than skewed, one quality associated with buffering.

Likewise, 19 of the 26 participants had more responses with an SPE of -1 than could be explained by identification errors. The binomial test for these participants yielded p-values ranging from 1.4×10^{-13} to .02. This suggests that these participants responded with letters represented in a buffer, because their reports of stimuli from before the cue were more frequent than expected under identification failures. The remaining participants had p-values ranging from .12 to .62.

The results of this analysis indicates that participants in Goodbourn and Holcombe (2015) did indeed produce SPE distributions likely reflecting buffered information. The distributions were largely symmetric, with the exception of one participant, and the majority of participants had more SPEs of -1 than could be accounted for by a liberal estimate of the incidence of identification failures.

2.4 Experiment 1: Many RSVP Rates

We believe that the mixture modelling procedure can recover the temporal properties of efficacious selections while accounting for those responses that result from identification failures. To test this, we employ a manipulation that increases identification errors and thus should result in changes in the efficacy estimates but no changes in the temporal parameters. Changes in efficacy could result in changes in the temporal estimates of the model (latency and precision) if responses that are identification errors are modelled as efficacious by the model fitting procedure. To affect efficacy, we alter the rate of presentation.

This chapter also constitutes a test that the latency and precision values estimated by the mixture modelling procedure are temporal, rather than item based. One theory of attentional selection in RSVP states that all stimuli are processed to a conceptual level, and that attention operates on these item-based representations (Chun & Potter, 1995). That is, the deployment of attention occurs on a scale defined by items (and thus serial positions) rather than being truly temporal. In this case, changing the presentation rate should result in selection of the item at the same serial position across different presentation rates. If however, selection is temporal rather than based on item serial position, changing the presentation rate should result in selection at a constant time across the different rates.

Manipulating the rate of presentation should affect the rate of identification errors, but the temporal properties of selection should not differ. Bowman and Wyble (2007) manipulated the rate of presentation in an RSVP task with two targets to investigate the time course of the attentional blink, a reduction in the probability of reporting the second of two RSVP targets if the first target was presented 200 to 500 ms prior and successfully reported. They presented stimuli at SOAs of 54 ms or 94 ms and found that the millisecond time course of the blink, which is thought to result from changes in the ability to deploy attention over time, was steady across presentation rates. However the accuracy of selection was poorer for the faster presentation rate. Vul, Hanus and Kanwisher (2008) found that correct identifications of the first target in an attentional blink paradigm are less frequent at a fast presentation rate (60 ms/item) relative to a slower rate (120 ms/item), but the center of mass of the temporal distribution of reports was the same. The Vul, Hanus and Kanwisher (2008) measure of the centre of mass is based on measuring the probability of reports within a fixed temporal window around the target. However, this analysis - which does not control for identification failures - has been shown to produce different estimates of the temporal qualities of selection in the AB for the second target, relative to mixture modelling (Goodbourn et al., 2016).

Thus, we expect that manipulating the presentation rate of RSVP streams should result in decreases in efficacy as the presentation rate increased. There should be no change in latency or precision of selection, because these variables correspond to the timing properties of visual selection and should be constant regardless of the probability of an efficacious selection.

Temporal, rather than item-based, selection in RSVP also could help answer an effect observed by Botella and Eriksen (1991) regarding the symmetry of SPE distributions. The presence of symmetric distributions of SPEs around the target has been interpreted as evidence for parallel processing of the cue and target (Botella & Eriksen, 1992; Broadbent & Broadbent, 1986; Gathercole & Broadbent, 1984; McLean et al., 1983). However, Botella and Eriksen (1991) observed that the pattern of SPEs for selections from a single stream shifted from symmetric around the target to post-target with increases in presentation rate. This according to the literature on parallel processing, should be interpreted as a shift in strategy, but participants did not know the presentation rate prior to each trial and thus could not have changed their strategy accordingly.

Botella and Eriksen's (1991) result suggests that we may observe skew when the presentation rate increases. They did not propose a reason for their effect. It may be that slow presentation rates provide coarse temporal information, which obscures skew in the data. However, Botella and Eriksen (1991) analysed their data in terms of SPEs rather than milliseconds and this may explain why they interpreted their results as demonstrating evidence for skew. They considered there to be evidence for a symmetric distribution when the proportion of SPEs from before the target was not significantly different from the proportion from after the target. However, their results are consistent with a mean latency of selection in time, rather than in SPEs, that is positive and does not change with the presentation rate. For slow presentation rates, the latency of selection may coincide with the target. As presentation rates increase, attentional selection with an unchanged temporal latency may select a later item. This would lead to a greater proportion of post-target reports than pre-target reports with faster presentation rates - a phenomenon that Botella and Eriksen (1991) interpret as evidence for skew. We compare model fits for the skewed gamma mixture and the symmetric Gaussian mixture and analyse our latency estimates in milliseconds rather than SPEs to assess whether there is evidence for a change in symmetry as the presentation rate increases.

2.4.1 Method

Participants. The data were collected by Patrick Goodbourn in 2012, and comprised six psychophysically-experienced participants including Patrick Goodbourn himself and Alex O. Holcombe.

Stimuli. Participants viewed two streams consisting of white 4º high Menlo letters on a black background. The streams were presented above a central fixation point (a 0.125º radius white circle) at eccentricities of 6.0º with a center-to-center distance of 6.0º between streams. The cues were 4.5º diameter white circles with line widths of 0.1º centered on the streams. The streams were random samples of the alphabet with no repeats and the letters V and N removed.

Stimuli were presented at one of four rates: 6, 8, 12 or 24 items/second, corresponding to SOAs of 166.67, 125.00, 83.33 and 41.67 ms, respectively. Each stimulus was presented for two-thirds of the SOA for that trial, with a one-third SOA blank period between stimuli.

Procedure. The program instructed participants to maintain their gaze on the fixation point and report the letters that were cued by the white ring on each trial. The onset of the ring cues were always simultaneous and their serial position on each trial was a sample from a uniform distribution bounded at 7 and 18.

After the streams had completed for a particular trial, participants saw the stimuli presented on each trial vertically arrayed on the left and right of fixation in alphabetical order. These were response arrays and participants clicked on the letter they wished to report for that trial. One side was randomly chosen to query first by presenting it in high contrast, while the array on the opposite side was presented in low contrast and could not be clicked on.

Design. Trials with a particular rate were blocked in sets of 100 trials. Participants completed one block of each rate and performed the blocks in a random order.

2.4.2 Results

Model Comparison. There was very little change in the model favoured by the data as the presentation rate increased. Most participants produced data that were favoured by the Gaussian mixture model. More of the fits were ambiguous when stimuli were presented at a rate of six items/second than in faster conditions.

Table 6

The number of fits favouring each type of model by rate and stream. “Neither” refers to a Bayes factor between .33 and 3.

Rate	Stream	Gaussian	Gamma	Neither
6 Items/second	Left	3	0	3
6 Items/second	Right	2	1	3
8 Items/second	Left	5	0	1
8 Items/second	Right	4	0	2
12 Items/second	Left	5	0	1

12 Items/second	Right	4	1	1
24 Items/second	Left	4	0	2
24 Items/second	Right	4	1	1

Parameter Estimates

Efficacy. The descriptive statistics for the estimates of efficacy, the probability of a response informed by the cue, from the Gaussian model are presented in Table **7** and Figure **8**. To investigate whether efficacy decreased as rate of presentation increased, we fit a series of linear models using the BayesFactor package in R (Morey & Rouder, 2018). These models estimated the relationship between efficacy and stream (left or right) and rate, with random intercepts by participant. After fitting the models, we computed inclusion Bayes factors for each independent variable. An inclusion Bayes factor for a particular variable compares the likelihood of the observations under models including that variable to the likelihood under those models in which the variable was absent, a form of Bayesian model averaging (Hinne, Gronau, van den Bergh, & Wagenmakers, 2019). The inclusion Bayes factor is a ratio of these likelihoods, where values greater than one indicate evidence for a particular variable, and values less than one indicate evidence against it.

The inclusion BFs favour a difference in efficacy between streams, the left stream has a higher efficacy than the right (BF_{10} = 9.45 x 10^{10}) - this is Goodbourn and Holcombe's (2015) pseudoextinction effect. Likewise, There is a decrease in efficacy as the rate of presentation increases (BF_{10} = 10047.9). These factors do not interact (BF_{10} = 0.18). The Bayes factor for their interaction is less than 1, indicating evidence for the null.

Table 7

Descriptive statistics for efficacy by rate and stream position

Rate	Left Mean (SD)	Right Mean (SD)

6 Items/Second	0.92 (0.03)	0.77 (0.11)
8 Items/Second	0.86(0.03)	0.72 (0.18)
12 Items/Second	0.78 (0.07)	0.62 (0.07)
24 Items/Second	0.53 (0.11)	0.27 (0.18)

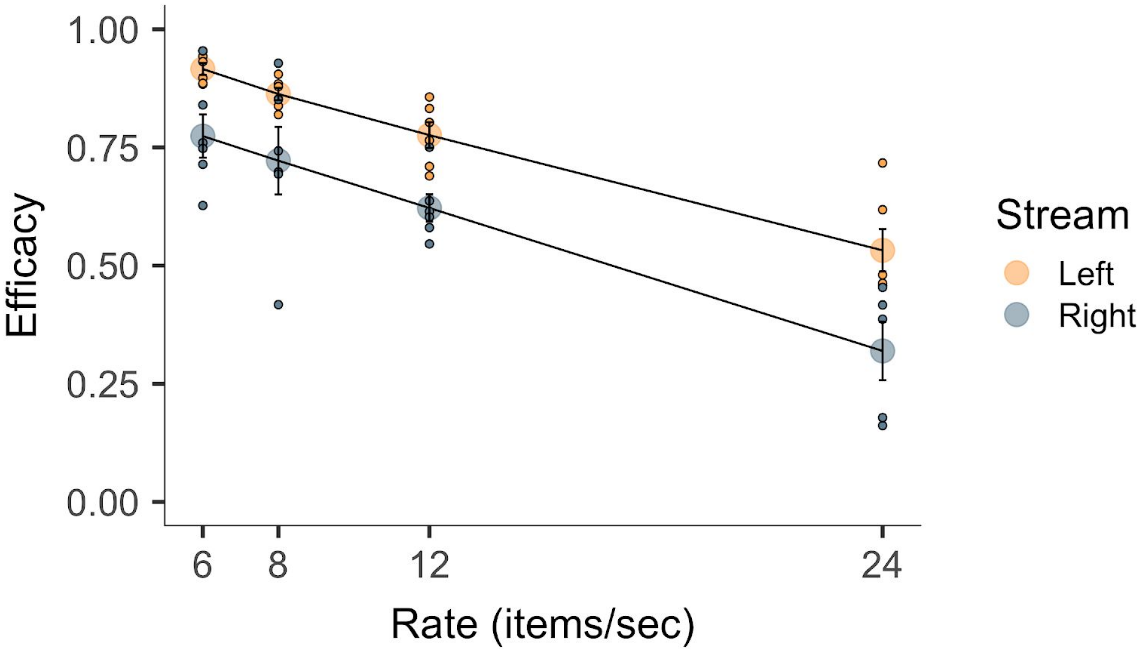

Figure 8. Efficacy estimates for the different rate conditions. The small points represent the data. The larger points with error bars are the means±SE. Efficacy decreases with faster presentation rates and the left stream is more efficacious than the right, but these factors do not interact.

Latency. The data for Latency, the mean timing of efficacious selections relative to the onset of the cue, were consistent with the null hypothesis of no change as the rate of presentation increased. Descriptive statistics for the latency estimates are presented in Table 8 and the data are presented in Figure 9. We fit the same kind of models as those described in the efficacy analyses, but this time latency, in milliseconds was the dependent variable. Inclusion Bayes factors for the predictors in the models indicated that the data favoured no effect for rate (BF_{10} = 0.43). or the interaction between rate and stream (BF_{10} =

0.45). However the inclusion Bayes factor for the effect of stream is close to one (BF_{10} = 1.21), indicating that the present data do not allow us to distinguish between the null and alternative hypotheses for this effect.

Table 8

Descriptive statistics for the latency estimates in milliseconds

Rate	Left Mean (SD)	Right Mean (SD)
6 Items/Second	25.54 (35.90)	46.13 (38.96)
8 Items/Second	39.78 (32.40)	52.41 (24.65)
12 Items/Second	45.39 (32.28)	67.32 (73.54)
24 Items/Second	56.64 (46.91)	58.31 (53.32)

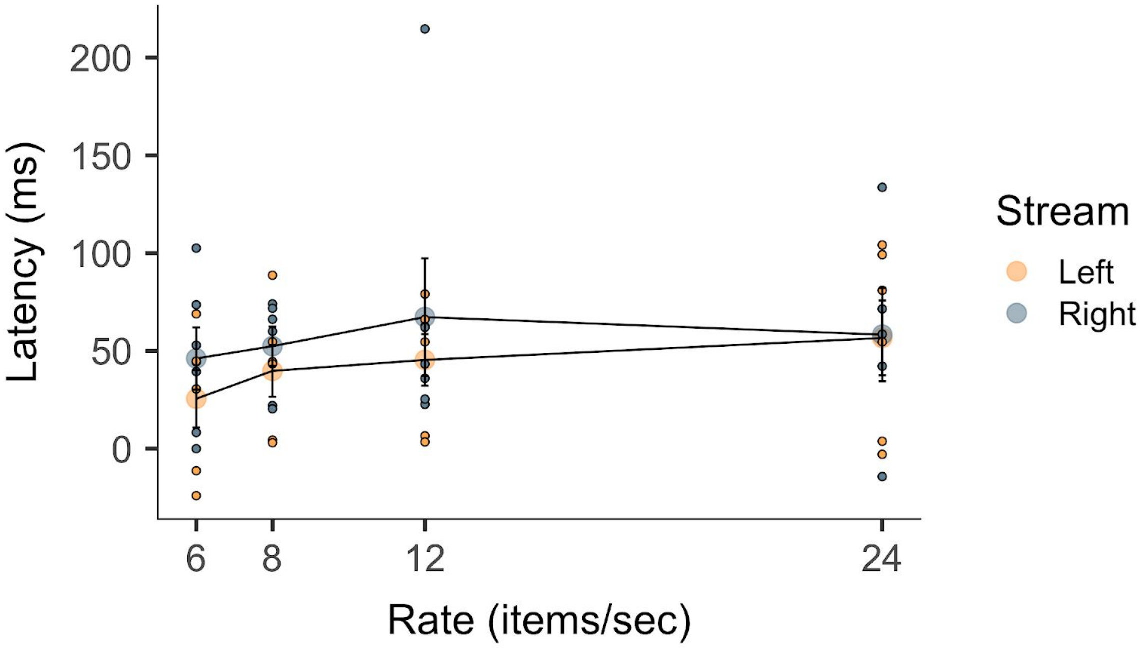

Figure 9. Latency estimates for the different rates and streams. Latency does not vary as the rate of presentation increases.

Precision. Precision estimates are presented in Table 9 and Figure 10. Inclusion BFs demonstrated that the data were consistent with no change in the rate of presentation

(BF_{10} = 0.37) and no effect of the position of the stream (BF_{10} = 0.29). However, the Bayes factor for the interaction between these variables is close to 1 (BF_{10} = .85), so we cannot draw any conclusions about the interaction between rate and stream in this experiment.

Table 9

Descriptive statistics for the precision estimates in milliseconds

Rate	Left Mean (SD)	Right Mean (SD)
6 Items/Second	53.6 (8.88)	61.15 (41.12)
8 Items/Second	60.67 (10.38)	60.88 (9.67)
12 Items/Second	65.13 (19.35)	94.02 (77.27)
24 Items/Second	82.22 (28.58)	51.19 (12.97)

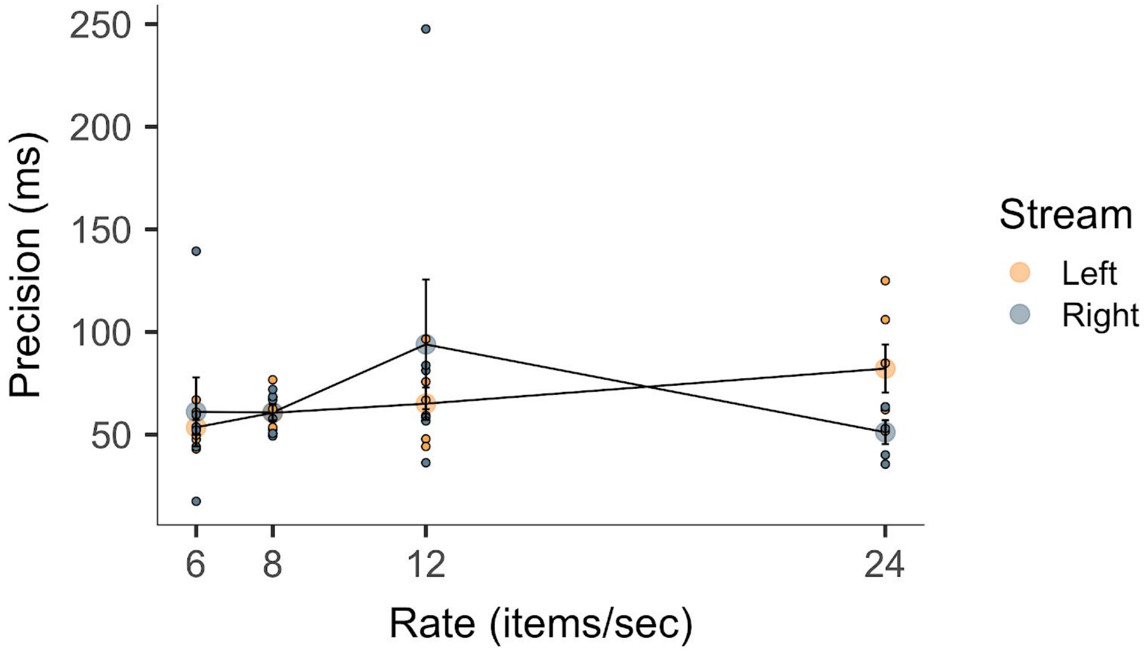

Figure 10. Precision estimates for the different rates and streams. The large points and error bars are means ±SE. The small points are the individual estimates.

2.4.3. Discussion

As expected, efficacy decreased as the rate of presentation increased, but latency and precision - estimates of the temporal properties of selection in this task - did not change. Likewise, the extent to which the Gaussian mixture model was favoured by participants' data did not depend on the rate of presentation in any systematic fashion.

The results of this study confirm that the latency and precision estimates of the mixture model reflect the temporal dynamics of selection rather than the selection of stimuli based on their serial position, as might be predicted by Chun and Potter (1995). Selection occurs based on time. Estimates of latency, the mean time of efficacious selections, were more likely under the null hypothesis than the alternative when compared across presentation rates. The temporal estimates were also steady as the rate of efficacious selections changed, demonstrating that the mixture modelling procedure can distinguish between efficacious reports and identification errors.

We also demonstrated that for the participants whose data were favoured by either the Gaussian or the gamma mixtures, there was not more evidence for skewed SPE distributions as the presentation rate increased. This is inconsistent with Botella and Eriksen's (1991) observations. However we analyse our data in terms of time (milliseconds) rather than SPEs and observe that the latency of efficacious selections is steady across presentation rates. This suggests that Botella and Eriksen's (1991) data do not indicate changes in the shape of the SPE distribution, but instead represent the fact that if the latency of selection is steady and positive, increasing the presentation rate will result in the selection of items with a greater SPE.

2.5 Summary

In this chapter, we outline a mixture modelling procedure for identifying the different temporal distributions of responses predicted by buffering and attention shifts. We also outlined a procedure for testing the frequency of a particular SPE, relative to the frequency expected when participants make identification failures. These tests reveal evidence for buffering in data from Goodbourn and Holcombe (2015). SPEs of -1 in these data were more frequent than expected based on identification failures, and the shape of the efficacious distribution was symmetric - both qualities associated with buffering. We also show that the latency and precision estimates of the mixture model are temporal, because they are robust to changes in efficacy.

Chapter 3: Endogenous and Exogenous Cues

The literature on attentional effects in vision distinguishes between different sources of attentional orienting. Attentional orienting can be elicited reflexively by a stimulus or voluntarily directed by an observer. One difference between these processes for deploying attention - named exogenous and endogenous attention, respectively - is their timing. Exogenous attention is often observed to be faster than endogenous attention. In Chapter 2, we demonstrated that the mixture modelling procedure reliably estimates the temporal qualities of selection in RSVP. In the present chapter, we investigate endogenous and exogenous attention shifts in RSVP and use mixture modelling to assess the efficacy, latency and precision of selection in response to such cues.

Comparisons between endogenous and exogenous attention in RSVP are rare, we have found no statistical comparisons. The majority of investigations of the time course of endogenous attention shifts have been cue-target lag studies, where the time between a cue and a static stimulus is varied (Cheal & Lyon, 1991; Liu et al., 2007; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989). This has provided valuable information about the time course of attentional enhancement. However, as we have noted, cue-target lag studies provide cumulative distributions of attentional timing. RSVP may also reveal phenomena not observed with cue-target lag. Goodbourn and Holcombe (2015) observed efficacious selections from RSVP streams that included items from before the cue - a phenomenon that cue-target lag studies do not observe.

Endogenous shifts of attention are named such because they are believed to be voluntary. To elicit an endogenous shift, researchers often train participants to interpret a central symbolic cue - such as an arrow - that indicates the location of an upcoming peripheral target stimulus with some probability. To attend to the target stimulus, the observer must interpret this stimulus, which provides only symbolic information about the target's location, and voluntarily shift their attention to the cued location. When a symbolic cue indicates the location of a target stimulus, responses are faster and more accurate than when the cue indicates the wrong location or the cue is uninformative about the target's location (Cheal & Lyon, 1991; Cheal, Lyon, & Hubbard, 1991; Eriksen & Yeh, 1985; Giordano et al., 2009; Ling & Carrasco, 2006; Müller & Rabbitt, 1989; Posner, 1980).

Key evidence for the voluntary nature of shifts in response to symbolic cues comes from manipulations of the probability that the cue indicates the location of a target stimulus, known as cue validity. Participants appear to allocate endogenously oriented attention based on the validity of the cue. More valid cues lead to faster processing at the cued location relative to less valid cues, as demonstrated by reaction times (Eriksen & Yeh, 1985; Madden, 1992; Vossel, Thiel, & Fink, 2006) and speed-accuracy tradeoff analyses (Giordano et al., 2009). Increasing validity is also associated with greater costs if the target appears at an uncued location (Eriksen & Yeh, 1985; Giordano et al., 2009; Vossel et al., 2006).

Exogenous shifts of attention indicate no such effects of validity. Instead, they appear to be reflexive. To elicit an exogenous shift of attention, a rapidly appearing (Remington et al., 1992; Yantis & Jonides, 1984) stimulus is presented at the location of the target stimulus. This stimulus yields, like endogenous attention, faster reaction times and improved accuracy at the cued location (Cheal & Lyon, 1991; Cheal et al., 1991; Ling & Carrasco, 2006; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Remington et al., 1992). However, unlike endogenous cueing, the exogenous cueing effect does not appear to be under participants' control. Nakayama and Mackeben (1989) demonstrated that the cue-target lag of an exogenous cue still affected visual search performance when participants already knew the location of the target. Manipulating exogenous cue validity produces no change in the effect of the cue on the speed of information accrual or signal sensitivity (Giordano et al., 2009).

Even when exogenous cues are uninformative about the location of a target stimulus (50% validity), cueing a stimulus location with an exogenous cue still causes improvements in contrast thresholds (Pestilli & Carrasco, 2005), acuity (Montagna, Pestilli, & Carrasco, 2009), and the speed and accuracy of orientation judgements (Liu, Pestilli, & Carrasco, 2005). This advantage is reflected in the larger magnitude of the fMRI response to cued stimuli with an uninformative cue (Liu et al., 2005). These results indicate that exogenous cueing is reflexive, participants cannot ignore exogenous cues based on the probability that they are informative.

Much of the literature on exogenous and endogenous attention has concentrated on their speed, measured by changes in performance with different lags between a cue and a subsequently presented static stimulus, such as a visual search array or Gabor patch. Endogenous cueing effects require a longer cue-target lag than exogenous cues. For endogenous cues, participants' accuracy reaches its maximum when a valid cue is presented around 300 ms before the target stimulus (Cheal & Lyon, 1991; Liu et al., 2007; Müller & Rabbitt, 1989). Exogenous attention, on the other hand, seems most effective when the cue precedes the target stimulus by around 120 ms (Cheal & Lyon, 1991; Folk et al., 1992; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Posner & Cohen, 1984; Remington et al., 1992).

The cue-target lag studies, in which the cue precedes the target stimulus by some lag, hints at the distribution of attentional arrival times. However, as discussed in Chapter 1, changes in performance with different cue-target lags tell us only that attention arrived at the cued location at or before the target on some set of trials. This is further complicated by exogenous attention's transient nature. Accuracy reduces after the peak at around 120 ms as cue-target lags increase, indicating that attention arrives at the cued location and then disengages (Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989).

RSVP studies, in which participants view a sequence of stimuli and select one for report based on some targeted feature, yield more insight into the timing of attention because the time at which a reported item was presented likely reflects the timing of attention on that trial. Weichselgartner and Sperling (1987) assessed the timing of exogenous selection from a single RSVP stream in which the cue was the presentation of a brightened letter or the appearance of a square around the stream. Participants in this task reported four stimuli from the stream, but the earliest reported stimulus can be taken as a measure of the attentional arrival time. The earliest reports tended to come from the cued stimulus, suggesting an attentional latency that was between 0 ms and 80 ms. Reeves and Sperling (1986) assessed selection latencies in a similar paradigm, but with an endogenous cue. In their task, participants monitored one peripheral RSVP stream for the onset of a target, then shifted their attention to another stream and reported four stimuli. This resulted in selection latencies of around 300ms, a result replicated by a control condition in Weichselgartner and Sperling (1987). Unfortunately, Weichselgartner and Sperling (1987) do not report statistical tests comparing exogenous and endogenous attention shifts.

These experiments provided insight into attentional selection latencies in RSVP, but they did not separate out the contribution of identification failures to the SPE distribution. Identification failures, responses where participants misidentified a selected letter or made a guess because they did not detect the cue, contaminate the SPE distribution because they result in responses of items from an almost-uniform range of times relative to the cue (Goodbourn & Holcombe, 2015). In the current experiments we used RSVP and mixture modelling to account for the contribution of identification errors. In doing so we can also estimate the temporal properties of endogenous and exogenous selection, as well as the efficacy of the different cues.

Investigations of the way in which endogenous attention can result in temporal selection in RSVP are rare. We have already mentioned Reeves and Sperling (1986) and Weichselgartner and Sperling (1987), in which participants reported several stimuli after an exogenous or endogenous attention shift. However because the SPE distributions in these tasks were not mixture modelled, it is unclear what the efficacy, latency and precision of selection were under these conditions. In an RSVP task with a central cue, Du and Abrams (2010) used an investigated voluntary attentional orienting during the attentional blink, but only measured participants' accuracy for reporting the second of two RSVP targets located in the same stream and did not report temporal measures of selection.

In this chapter, we investigate the temporal properties and efficacy of endogenous and exogenous cued selection from one of several RSVP streams. Such an investigation matters because the majority of RSVP research uses exogenous cues such as a cue that surrounds the target (i.e. Goodbourn & Holcombe, 2015; Vul & Rich, 2010; Weichselgartner & Sperling, 1987) or does not model the SPE distributions from endogenous selection in a manner that accounts for efficacy. Furthermore, investigations of endogenous attention's temporal property typically use a static target stimulus - like a visual search array - and manipulate cue-target lag such that the time of attentional selection on a particular trial is unclear (i.e. Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989).

3.1 Method

Participants: Participants were seven psychophysically-experienced observers (five graduate students, one professor, and one undergraduate), including the author of this chapter and his supervisor. Due to a computer error, data for one participant in the endogenous condition were lost. This participant was excluded from the analysis, leaving six participants with complete data. The sample size was not based on power calculations or a Bayesian stopping rule, and was instead determined by the availability of participants.

Stimuli: Stimuli were 26 Arial letters, the tallest of which were 3º of visual angle high. There were six simultaneous RSVP streams in this experiment. Each stream was made up of a random sequence of the letters of the alphabet, with no repeats. The order of items within a stream was a random shuffle on each trial. The streams were presented equally spaced around an imaginary circle, centered on a fixation, with a radius of 6º. The fixation point was also a circle and subtended 0.62º.

The streams were presented at a rate of approximately 15Hz. Every stream appeared and was updated synchronously with the others. The monitor refreshed at a rate of 60Hz. Each letter in the stream was presented for three monitor refreshes (50 ms), followed by a blank of one monitor refresh (16.67 ms) before the onset of the next stimulus. The target item was indicated by the onset of a cue (described below), which was simultaneous with the onset of the target item and indicated its position.

In endogenous trials, the cue was a red pixel (RGB: 255,0,0), 0.03º in diameter, presented 0.07º from the center of the fixation point in the direction of the stream containing the target. Its duration was the same as a frame of the stimuli, 50 ms.

In exogenous trials, the cue was a circle with a diameter of 5º and a line-width of .07º presented centred on the target letter so that it surrounded it. Its duration was the same as a frame of the stimuli, 50 ms.

Apparatus: Participants viewed the stimuli from a distance of 57 cm in a darkened room. Stimuli were presented on a Mitsubishi Diamond Pro 2070SB CRT monitor at a resolution of 1024 x 768. The width of the monitor was 40.5 cm and its refresh rate was 60 Hz. The central fixation point was presented at the centre of the screen, 512 pixels below and 384 pixels to the right of the top left corner of the screen. This experiment was not eyetracked

Procedure: Participants were told to maintain fixation on the point at the center of the screen and report the cued item on each trial. They were instructed that the cue designating the target would either be a circle around the letter, or a small point at fixation offset in the direction of the target item. On each trial, participants first saw the fixation point. After a random interval between 300 ms and 800 ms, the RSVP streams appeared and began the sequence. The cue appeared at a random interval of 7 to 11 serial positions (467 to 733ms) from the start of the trial. The cued stream was randomly chosen on each trial.

At the end of each trial, the alphabet was horizontally arrayed on the screen, and the participant had to select the cued letter with the mouse.

Design: There were two conditions in this experiment: endogenous and exogenous. Each participant was presented with both conditions, which were blocked. Each condition consisted of 200 trials. Each of the possible temporal positions of the cue occurred 40 times in each condition, with the trial order of temporal positions random.

Analysis: We fit two mixture models, a gamma mixed with a windowed-uniform model and a Gaussian mixed with a windowed-uniform model, to each participant's SPE data in each condition. The first 20 trials were not included in the analyses because the early trials are likely to reflect changes in participants' ability to perform the task. The model fitting was performed in R (R Core Team, 2019) with maximum-likelihood estimation using a custom package (https://doi.org/10.5281/zenodo.3545085).

We compared model fits for each set of SPEs using Bayes factors estimated from the Bayesian Information Criterion for each model (Wagenmakers, 2007 Eq. 10). The Bayesian Information Criterion allows us to estimate the Bayes factor with an uninformative prior over the model parameters (Raftery, 1999; Wagenmakers, 2007). The Bayes factor provides the ratio of the weight of the evidence for each model in the data and thus can be used for model comparison. We will refer to Bayes factors that were greater than 3:1 favouring either model as providing evidence for one of the models. Those that do not exceed 3:1 in favour of either model we will describe as providing equivocal evidence.

We compared the parameter estimates between conditions using paired Bayesian t-tests with a JSZ prior on the effect size (Rouder, Speckman, Sun, Morey, & Iverson, 2009) using the BayesFactor package (Morey & Rouder, 2018) in .

3.2 Results

Model Fits. For the exogenous cue, four of the six participants produced SPE distributions that were best fit by the symmetric, Gaussian mixture model. One participant in this condition produced data that were best fit by the skewed, Gamma model, and one participant produced data that did not favour either model.

For the endogenous cue, three participants produced data that were best fit by the Gaussian model, two participants were best fit by the Gamma model, and one participant's data did not favour either model.

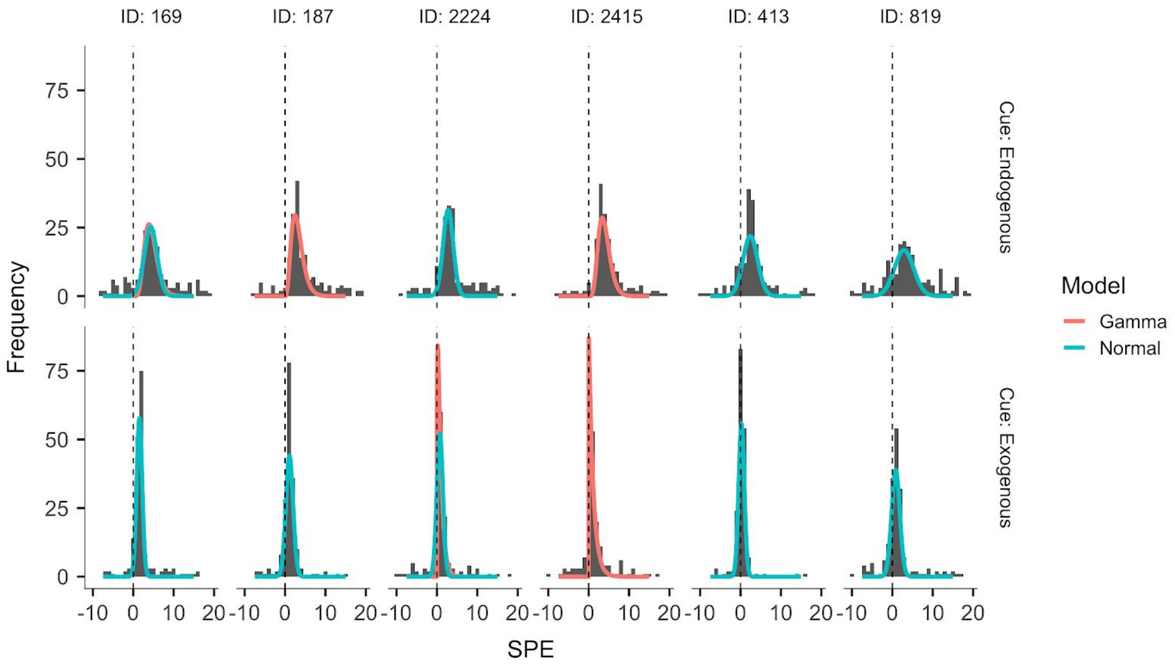

Figure 1. SPE histograms for the six participants who had data from both conditions. The best-fitting efficacious distribution is shown, scaled to near the height of the histograms. In instances where the data did not favour either model, we plot both efficacious distributions.

Parameter estimates. We report the parameter estimates for the Gaussian model here, but the direction of the differences in estimates between conditions do not change when the estimates of the Gamma distribution are reported.

Efficacy. Efficacy, the probability of making a response informed by the successful identification of the cue (as captured by the gamma or Gaussian distribution), was substantially lower for the endogenous condition (M = .61, SD = .10) than for the exogenous condition (M = .82, SD = .07, BF_{10} = 19.08)

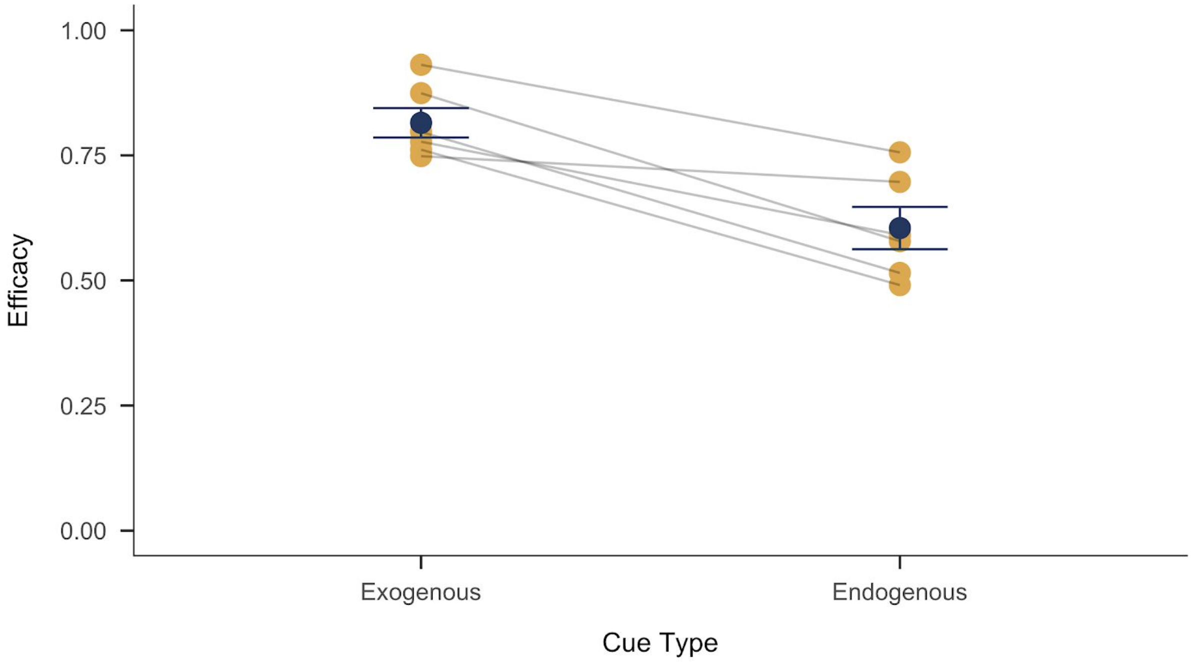

Figure 2. Efficacy estimates for the different kinds of cue. The yellow points represent individual participants' estimates. The dark points and error bars represent mean ± SE. Participants are less likely to make an efficacious selection with an endogenous cue than an exogenous cue.

Latency. The mean time of a reported efficacious item was much later when participants selected items in response to an endogenous cue (M = 212.0 ms, SD = 47.6 ms) than with an exogenous cue (M = 58.2 ms, SD=28.2 ms, BF_{10} = 334.90)

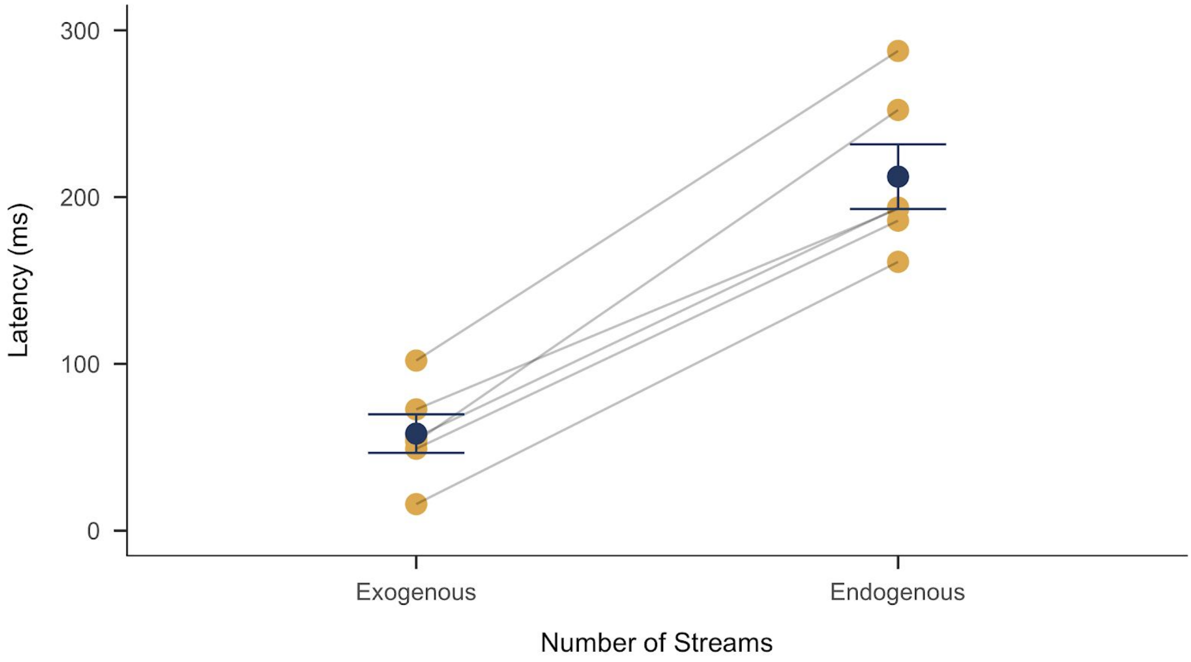

Figure 3. Latency estimates for the different kinds of cue. Efficacious responses come from a later point in time, relative to the cue's onset, with an endogenous cue than with an exogenous cue. The yellow points represent individual participants' estimates. The blue points and error bars represent mean ± SE.

Precision. The distribution of efficacious selections was broader in the endogenous condition – precision, the standard deviation of efficacious selections, was greater (M = 109.0 ms, SD = 28.4 ms) than for the exogenous condition (M = 56.2 ms, SD = 9.3 ms, BF_{10} = 12.5)

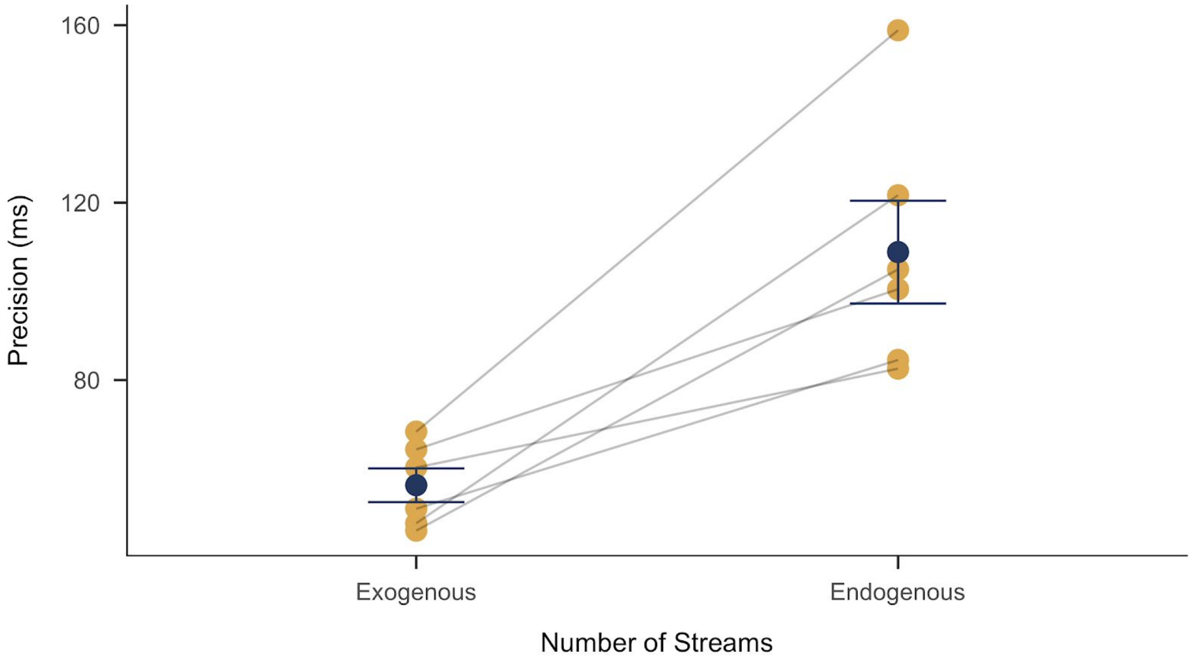

Figure 4. Precision estimates for the two cues. Efficacious responses are less variable with an exogenous cue than with an endogenous cue. The yellow points represent individual participants' estimates. The dark points and error bars represent mean ± SE.

Binomial Analysis. We used the binomial analysis to assess the extent to which either cue led to efficacious selections of items presented prior to the target. In the exogenous condition, two participants produced more selections of items with an SPE of -1 than could be explained by non-efficacious responses, $p = 5.6 \times 10^{-13}$ and .007. The same two participants had more selections with an SPE of -1 than expected by non-efficacious responding in the endogenous condition as well, p = .01 and .049.

3.3 Discussion

Exogenous cueing of an RSVP stream is more likely to result in an efficacious selection than endogenous cueing. When participants make an efficacious selection, they are faster and less variable in time with an exogenous cue than an endogenous cue. The data do not indicate not much change in the shape of the efficacious distribution when participants select an item with an endogenous cue. Only one more participant demonstrated a skewed efficacious distribution with an endogenous cue than an exogenous.

This is the first comparison of the latency and precision of endogenous and exogenous cueing in RSVP with a procedure that accounts for failures to detect the cue or identify a relevant stimulus. As we predicted, the latency of endogenous cueing was greater than that of exogenous cueing, consistent with previous observations that endogenous attention takes longer to orient to a cued location than does exogenous attention (Cheal & Lyon, 1991; Müller & Rabbitt, 1989).

Endogenous cues are less likely to result in an efficacious selection than exogenous cues. This effect mirrors the observations by Müller and Rabbitt (1989) and Cheal and Lyon (1991) that exogenous cues lead to a higher peak accuracy than endogenous cues when locating a target. We believe that the reflexive nature of exogenous attention shifts and the voluntary nature of endogenous shifts explain this. Upon detection of the exogenous cue, participants involuntarily shift attention to the cued location (Giordano et al., 2009; Müller & Rabbitt, 1989; Pestilli & Carrasco, 2005). On the other hand, the detection of an endogenous cue is not a sufficient condition for an endogenous attention shift. Participants must interpret the cue - in this experiment a judgement about where the line is pointing - and make a voluntary shift of attention to the cued location. These additional steps may fail, resulting in the reduction in efficacy observed in this experiment.

The attentional latency estimates reported here are smaller than those reported in a test of exogenously- and endogenously-cued attention with many simultaneous dynamic stimuli. The participants in an experiment by Carlson, Hogendoorn and Verstraten (2006) viewed 10 clock faces arranged in a circle, much like our six RSVP streams, each with a rotating hand. One of these was cued with an endogenous cue (a central line) or an exogenous cue (a change in colour), and participants reported the orientation of the hand at the time of the cue. Exogenous and endogenous cueing resulted in mean latencies of 140 ms and 250 ms respectively - almost double the latency estimates observed in our experiment. The increased latency of selection with the clock hands relative to RSVP streams is likely related to the continuous movement of the clock hands. When participants must select from a stream of temporally autocorrelated stimuli, like a smoothly changing clock hand, attentional selection latencies are longer than when those stimuli change randomly, as in RSVP (Callahan-Flintoft et al., 2019).

Precision, the variability of attentional selection in time, has not received much attention in the literature. Using the same clock stimuli as Carlson, Hogendoorn and Verstraten (2006), Hogendoorn et al. (2010) found that when participants had to monitor six simultaneous clocks for the onset of a cue, precision was approximately 120 ms (Hogendoorn et al, 2006, Figure 2). When participants make an identification error, their responses have a uniform distribution. This means that attempts to estimate precision without accounting for identification failures will overestimate precision, such that selections appear more variable than they are when identification failures are accounted for. This means that Hogendoorn et al.'s (2006) precision estimates are overestimates, because they represent variance in both identification failures and efficacious selection, as the authors note. Weichselgartner and Sperling (1987) briefly note that the timing of an endogenous shift of attention from one RSVP stream to another typically has a standard deviation of 100ms. This estimate also fails to account for efficacy and thus is an overestimate.

Here, accounting for identification failures, we find a wider precision - more temporal variance - with an endogenous cue than an exogenous cue. The wider variance of efficacious responses in the endogenous condition matters, because increasing the range of possible selection times means than endogenous selection more likely to result in a very delayed selection, one that misses a target in our task. The cueing stimulus within a condition was the same across trials, with minor differences in foveal location. Thus changes in precision estimates correspond to the effect of internal noise in the process of identifying the cue, interpreting it, and voluntarily shifting attention to the cued stream.

Noise is a fundamental property of neural information processing (Faisal, Selen, & Wolpert, 2008). It appears at all levels of the visual system from the absorption of photons by photoreceptors (Bialek, 1987), neurons' spiking rates (White, Rubinstein, & Kay, 2000), and sensory representations (Swets, Tanner, & Birdsall, 1961; Verghese, 2001). Much effort has been expended investigating noise in the visual system, because visual sensitivity depends not only on the physical properties of a stimulus, but the noise in visual processing as well (Pelli & Blakemore, 1990; Pelli & Farell, 1999). Investigating visual performance with a model that incorporates internal noise (Lu & Dosher, 1999) has provided valuable insight into changes in perceptual template associated with attention (Dosher & Lu, 2000a, 2000b; Lu & Dosher, 1998).

Our precision estimates indicate that there is more temporal noise in endogenously cued attentional orienting than exogenously cued orienting, but where in processing might this noise occur? In the exogenous cueing condition, the location of the target is given by the cue, because the cue surrounds the cued stream. Endogenous cues are symbolic, and the location they refer to must be interpreted, as we discussed above. This interpretation time is likely to be the source of the additional variance in endogenous orienting we observe here. It corresponds to the process of making a decision about the direction indicated by the cue (i.e. Palmer, Huk, & Shadlen, 2005), prior to attending to the stream located in that direction. We believe this because we can rule out spatial errors as contributing to precision differences between the conditions.

It is reasonable to expect more spatial errors with an endogenous cue than an exogenous. This is because the exogenous cue provides precise spatial information about the cued stream - the cue appears near the cued stream. The endogenous cue, on the other hand, requires interpretation in order for a participant to understand which stream is cued. This interpretation can fail, directing attention to the wrong stream, but this would result in a decrement in efficacy in our study, because the order of stimuli in the uncued streams is randomly distributed relative to the target in the cued stream. We observe such a decrement, so we do not rule out the possibility that this occurred in our experiment. However it is not the source of the precision changes because changes in efficacy do not affect precision estimates, as shown in Chapter 2 of this thesis. If misdirection of attention based on an ambiguous cue cannot explain the precision difference, then what remains is variance in the time taken to interpret the cue. Thus it is variability in the decision process related to the endogenous cue that we believe is likely to explain this increased variance.

Thus we propose that it is the interpretation component of endogenous attentional orienting that accounts for the decrease in efficacy and widening of precision in the endogenous condition relative to the exogenous condition, in which orienting is reflexive. One way in which to test the relationship between the precision of selection and cue interpretation is to instruct participants to make a motor response indicating the position indicated by the cue on a particular trial. Motor response times reflect a decision about the location of the cued stream - that is, cue interpretation - prior to any attempt to shift attention to that stream. Thus they should share variance with the attention shift to the cued location, which is informed by cue interpretation (Shih & Sperling, 2002). Changes in cue interpretation time, which should vary between different cues, will lead to different correlations between SPE and motor reaction time. Shih and Sperling (2002) found small but positive correlations between motor reaction time to an auditory cue and the timing of attentional selection in response to the same cue, suggesting that endogenous attentional orienting and the motor decision do share some component.

The correlations in Shih and Sperling (2002) include trials in which participants made identification failures. Ideally, we would correlate motor RTs and efficacious trials, but the mixture modelling procedure does not allow us to identify those trials that were efficacious and those that were not. Thus we would not be able to directly analyse the relationship between precision and motor RT variability. This would attenuate the correlation somewhat, because we would, on some trials, be associating motor RT and identification failures. However, such an analysis would still allow us to assess the relationship between interpretation and attentional selection’s variability.

Chapter 4: Buffering depends on the number of streams

The brain cannot fully process all the stimuli in a busy visual scene. Therefore, resources must be selectively allocated to the most behaviorally relevant stimuli. We refer to this ability as “attention”. In this chapter we investigate a phenomenon that is hard to reconcile with many theories of attentional selection RSVP - reports of items from before the time of a cue that cannot be explained by guessing or target misidentification. These reports occur when there are only a few possible streams that can be cued, but not when there are many.

Deploying attention in response to a cue takes time. Aspects of the time-course have been inferred from the timing of reported stimuli and changes in performance over cue-target lags. In studies that varied the time between when a cue was presented and a lone post-masked target, the effect of the cue-target interval on performance is interpreted as representing the dynamics of attentional allocation (Carrasco, 2011; Cheal & Lyon, 1991; Eriksen & Hoffman, 1973; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989).

In RSVP, stimuli are presented in rapid succession, and the time to complete an attention shift can be inferred from the temporal position of the stimulus reported by a participant on a particular trial (Reeves & Sperling, 1986). In a typical RSVP task, participants view a sequence of stimuli presented at a single spatial location and must report one item’s identity based on some target feature. For example, a stream of individual letters may be presented with a random letter cued by a change in color or an enclosing circle. If the sequence of stimuli contains no repeated letters, the time at which selection occurred can be inferred from the identity of the reported item. Over many trials, this method produces a distribution of selection errors that facilitates inferences about the timing of attentional selection under particular conditions (i.e Chun, 1997; Reeves & Sperling, 1986; Weichselgartner & Sperling, 1987). However, these inferences assume that any given response was not a lucky guess - an assumption that may not be true. We will return to this point later.

At least in some circumstances, a cue in an RSVP task appears to elicit reports of stimuli that seem inconsistent with a time-consuming shift of attention. A few researchers have suggested that the cue can result in participants selecting and report stimuli that were presented before the cue (Botella et al., 2001; Goodbourn & Holcombe, 2015; Holcombe et al., 2017). Such findings suggest that stimulus representations in those circumstances are maintained in a buffer so that they are still available when the cue is presented later.

Storage of irrelevant stimuli, also known as incidental memory, is not surprising in certain circumstances. Apart from a few papers, however, the literature on selection of a cued item from a stream of rapidly presented stimuli has not considered a role for memory of items presented before the cue. Already mentioned above were the papers that took the temporal distribution of performance as measuring the dynamics of a shift of attention– this assumes that stimuli reported are not drawn from memory. More recently, a large literature has developed around theories of the attentional blink and other RSVP tasks involving the report of multiple stimuli from a stream (Broadbent & Broadbent, 1987; Dux & Marois, 2009; Martens & Wyble, 2010; Reeves & Sperling, 1986; Shih & Sperling, 2002; Weichselgartner & Sperling, 1987). Yet very few have attempted to explain or model reports of stimuli from before the cue, and several assume that the quick succession of stimuli in RSVP causes masking, which prevents the selection of stimuli that are no longer presented (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002).

Goodbourn and Holcombe (2015) proposed the involvement, in RSVP tasks, of a buffer containing visual representations that were not entirely overwritten by the next letters, but rapidly decayed. Unlike an attention shift, which in RSVP is not assumed to operate on information stored in memory, the action of this buffer is not triggered by the cue. Instead, it is in operation before the cue appears, storing stimuli.

As a cue, Goodbourn & Holcombe (2015) presented a white ring around one of the letters in an RSVP stream. The resulting data were mixture modelled and the estimates of latency and precision suggested that the efficacious distribution included reports of items before the cue. Goodbourn & Holcombe suggested that such reports were evidence that the activation of letters prior to the cue can persist (buffering), and also that binding of the cue with an active letter representation is imprecise and sometimes results in temporal errors such as report of a still-active pre-cue letter. These pre-cue reports are the critical evidence for a buffering process. Such responses are unlikely under an attention shift, as that should be somewhat time-consuming, and thus they constitute critical evidence for buffering.

The buffer putatively responsible for reports of letters before the cue may be a brief store of visual information. Both iconic memory (Averbach & Coriell, 1961; Sperling, 1960) and the more recently-developed construct of fragile memory (Pinto, Sligte, Shapiro, & Lamme, 2013; Sligte et al., 2008) could explain this. However, these memory stores are thought to be overwritten by subsequent stimuli presented in the same location, which implies they will not outlast the appearance of the next letter in an RSVP stream. This assumption about masking has been included in several models of selection from RSVP (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002). Thus neither iconic nor fragile memory can explain Goodbourn and Holcombe’s suggestion that a cue in an RSVP stream of letters results in people sometimes reporting the letter before the cue.

4.1 Contributions to theory

We identify several models that attempted to explain performance in RSVP tasks below. However, for each of these models, either the presence of pre-cue reports or changes in the temporal distribution of reports constitutes phenomena that are relevant but unexplainable, or falsify assumptions about the persistence of representations.

One influential theory used a task of reporting multiple successive items from a stream after a cue (Reeves and Sperling, 1986). In this and some later theories (Grossberg & Stone, 1986; Shih & Sperling, 2002), information is buffered in iconic memory and attention acts as a gating mechanism through which this information can enter visual short term memory. This, in theory, allows for selections of items from before the cue if their representations persist long enough for the attention gating mechanism to open. However these models assume that the representation of one stimulus is overwritten by the appearance of another stimulus in the stream, consistent with the theory of iconic memory. Chun and Potter's (1995) theory of the attentional blink in RSVP shares this assumption. Thus, evidence for reports of pre-cue items that are efficacious falsifies these models' assumption that stimulus representations from RSVP are terminated by new stimulus.

Some published theories, however, do posit stimulus representations that persist during the processing of the next stimulus (Olivers & Meeter, 2008; Wyble, Bowman, & Nieuwenstein, 2009). These theories were designed to explain the attentional blink, and seem to have made little contact with literature on iconic memory and fragile memory. The persistence of representations despite subsequent stimuli in these models has the potential to explain pre-cue responses. However these models have no process that can result in errors in temporal selection, so they predict that there will be no reports of pre-cue stimuli, apart from those explainable by random guessing.

Work by Botella and colleagues directly addresses the possibility of reports of pre-cue stimuli. Botella, Suero & Barriopedro (2001) set out to explain the presence of pre- and post-cue responses in RSVP response distributions. In this model, following the ideas of Treisman & Gelade (1980), attention on some trials successfully focuses on the target, and in those cases there is no error. In the remaining trials, responses are made by a guess informed by buffered representations of unbound features from the items in the RSVP stream. The model's decay rate for non-target item representations falls from its peak to zero over a period of roughly 200-250 ms (Botella et al, 2001; Figure 6), which accommodates the occasional reports of pre-cue items based on incomplete feature information.

Vul and Rich's (2010) theory of how participants perform in cued RSVP tasks also implies buffered item features. Vul and Rich proposed that visual feature binding, including binding a cue to the simultaneous letter in a stream, involves uncertainty about which features occurred at the same time as the cue. Binding, on this theory, occurs via sampling from a representation of the probability distribution of which features were presented when. This theory implies a buffer, because if feature representations did not persist beyond stimulus presentation there would be no distribution.

In this chapter's experiments, we apply our binomial test to investigate whether pre-cue responses are efficacious. This is the first time that a statistical test has been applied to investigate the presence of buffered information, even though these responses are critical to the theories of Vul and Rich (2010), Botella et al (2001), and Goodbourn and Holcombe (2015).None of the published theories that assume a buffer appear to predict circumstances for when representations of items from before a single target will or will not be accessed. However, Wyble, Bowman, and Nieuwenstein (2009) theorize that during the attentional blink, selection is delayed, which might result in fewer pre-target reports, if their theory had a process for producing pre-target reports.

A further result unaccounted for by published theories is that not knowing which of many simultaneous streams will be cued has effects that include reducing the number of pre-cue reports. We will suggest that this results from delayed selection due to dilution of attention among the streams.

4.2 Critical Evidence for the buffer: Accounting for guessing

Reports of items from before the time of the cue are key evidence for buffering. At least some of these responses, however, will be identification failures - trials in which a participant does not detect the cue and makes a guess or misidentifies a selected letter. These failures, when aggregated across trials, result in a nearly-uniform distribution spanning all times relative to the cue (Goodbourn & Holcombe, 2015). The remaining trials are *efficacious* - those in which participants report a letter from a time related to the time of the cue as a result of processing that letter .

To investigate selection from RSVP, like others we use rapid presentation rates to avoid ceiling levels of performance. With a presentation rate of 12 items per second, Identification failures were estimated as comprising 25% of responses by Goodbourn & Holcombe (2015). Some of these identification failures will by chance be letters presented shortly before the cue. For example, when a participant makes a complete guess, occasionally that guess by chance will happen to be the letter presented just before the cue. In order to investigate buffering, we must account for trials in which identification failures led to a pre-cue report, because only efficacious pre-cue reports constitute evidence for buffering. The Goodbourn & Holcombe (2015) paradigm we use was designed to allow us to use mixture modelling to estimate of the proportion of trials that were identification failures and the proportion that were efficacious

4.3 Binomial Analysis of published data

To begin with, we assess the evidence for efficacious pre-cue selections in published RSVP datasets by applying our binomial test.

Vul et al. (2008) conducted an attentional blink experiment and implied that their data contained pre-cue efficacious responses, including for the first target. Taking their dataset, which was made available for the Goodbourn et al. (2016) re-analysis (https://osf.io/fs93m/), we applied the binomial test to quantitatively assess the evidence against the null hypothesis that the pre-cue responses were due to identification failures.

As an attentional blink experiment, Vul et al.'s (2008) task involved an RSVP stream of letters with a cue presented around two of them, separated by different numbers of items (lags). Figure 4 presents SPE distributions from participants for the first target (T1) for lags one to six from this experiment. Visual inspection suggests that the number of SPE = -1 responses exceeds that which could be explained by identification failures. We present the distributions for T1 only because items presented before the first target did not receive any cueing, whereas items from before the second target may have received cueing from the first target and are also subject to distortion from the attentional blink.

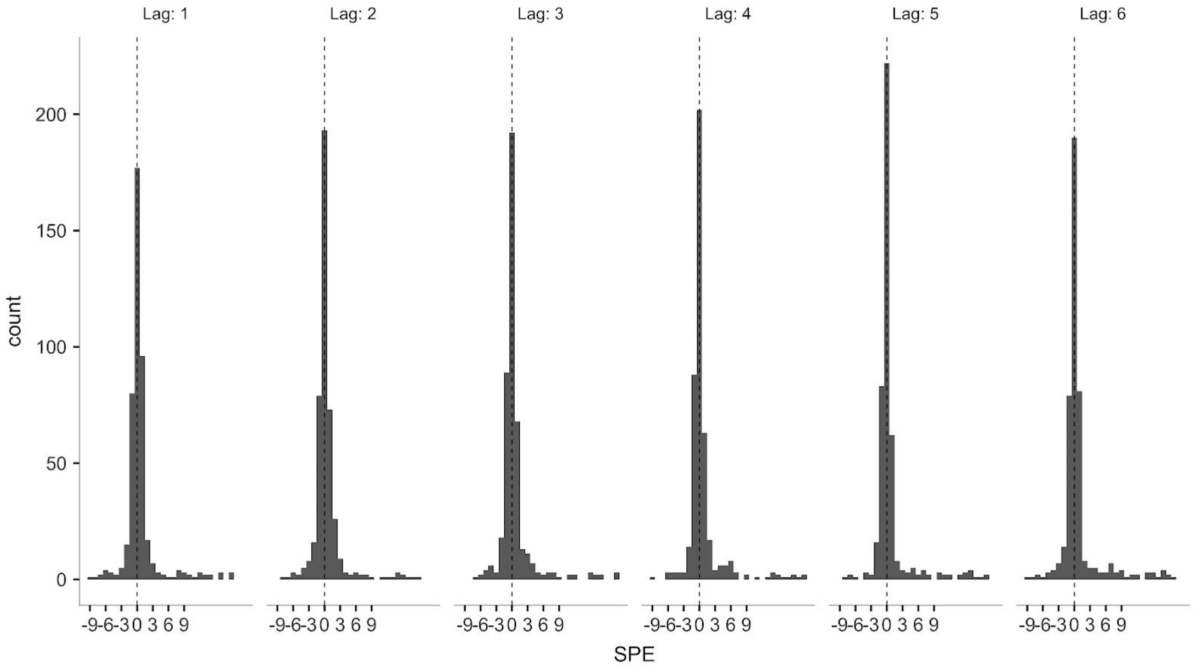

Figure 4. SPE histograms for T1 responses in Vul, Nieuwenstein and Kanwisher (2008), summed over participants. The dashed vertical line represents the target. Responses are approximately symmetric around the cue and the cluster around each distribution’s mode includes items from before the target, suggesting that some responses were drawn from a buffer.

For each lag between the first and second targets, we applied our binomial test to the data for the first target (T1). Using the $p < .05$ level, for nine or more of the eleven participants (depending upon the condition) we reject the null hypothesis that the SPE = -1 trials were generated entirely by identification failures (Table 1). *p*-values for the participants ranged from 3.9×10^{-18} to .047.

Table 1

Results of the binomial test applied to the data of the eleven participants in Vul et al. (2008)

Lag	Number of participants with p < .05
1	11
2	9
3	10
4	9

5	9
6	9
7	10
8	11
9	9
10	11

Goodbourn and Holcombe (2015) also claimed that for their own experiments, some of the SPE = -1 trials were a result of efficacious trials rather than misidentifications. We applied the binomial test to the data for their two-stream, single-target condition, in which participants responded to a single cue in one of two streams. Nineteen of the 26 participants had more responses with an SPE of -1 than could be explained by identification failures. *p*-values for these participants ranged from 1.4×10^{-13} to .03.

These results provide quantitative evidence for efficacious pre-cue responses in RSVP tasks - an indication of buffering despite the presence of masking in RSVP.

4.4 The present experiments

To investigate when the buffer is used, we tested various numbers of simultaneous RSVP streams, from two (the number used by Goodbourn and Holcombe, 2015) to 18. In each condition, we then assess the evidence for buffered letters using the binomial procedure. We also compare the fits of the Gaussian to the gamma distributions in mixture modelling to assess the evidence for attention shifts.

Initially, we thought of the manipulation of the number of simultaneous streams as an investigation of the capacity of buffering, because limited capacity restricts processing for many visual tasks (i.e. Hawkins, Houpt, Eidels, & Townsend, 2016; Luck & Vogel, 1997; Treisman & Gelade, 1980; Wolfe, 1998). This is explicit in the pre-registration for our first experiment (https://osf.io/7hkgd), where we assumed that increasing the number of streams would result in a change from selection from the buffer to attentional selection from the stream. However, we came to realise that the use of buffering could reflect the speed of attentional engagement as well. If a stimulus representation has not yet decayed during the presentation of the next item, then efficacious pre-cue responses may occur when attentional engagement is fast enough to select the persisting representation of an item presented in the very recent past. The size of the area that attention is dedicated to - in this case, a function of the number of RSVP streams - affects the speed of a reaction to a target. The larger the area, the slower the reaction time (Castiello & Umiltà, 1990, 1992). Thus pre-cue responses occurring when there are few RSVP streams may reflect the faster engagement of attention under those conditions.

The presence of efficacious pre-cue responses is critical evidence for the buffer. If we do not observe such responses using the binomial procedure, there is no evidence that buffered items are used to perform the RSVP task. We predict that as the number of streams increases, we will observe fewer efficacious responses from before the time of the cue, delays in the latency of selection and potentially more skew in the efficacious distributions.

Data and materials for all the experiments and analyses reported in this chapter can be found on Github (https://github.com/cludowici/RSVP_Dynamics/).

4.5 Experiment 1

4.5.1 Method

Participants. Ten University of Sydney first-year undergraduates (9 female) ranging in age from 18 - 23 years (*M* = 18.9 years, *SD* = 1.6) participated for credit for a psychology course. This experiment was pre-registered, including a data collection stopping rule of the Bayes factor for the effect of number of streams on latency exceeding 10:1 in favour of either the null or alternative hypothesis (https://osf.io/7hkgd). We did not create the custom prior referred to in the pre-registration because there were no existing effect size estimates.

Apparatus. The experiment was controlled by a Macbook Pro running Psychopy 1.85. Stimuli were presented on a Mitsubishi Diamond Pro 2070SB CRT screen with a resolution of 1024 x 768 pixels and a refresh rate of 60 Hz. Participants rested their head on a headrest 56.5 cm away from the screen. To enable exclusion of trials in which participants broke fixation, movement of the right eye was tracked with an SR Research Eyelink 1000.

Stimuli. The stimuli were white Sloan letters (Pelli, Robson, & Wilkins, 1988) that vertically subtended 0.9º. The cue was a 0.98º radius white ring with a line width of 0.07º. The fixation point was a white dot with a radius of 0.14º. On each trial, RSVP streams were presented simultaneously at two or eight positions. Each stream consisted of all letters of the alphabet except C and W, presented in a random order. The letters of each stream appeared at a rate of 15 items per second. Each letter was presented for three monitor frames (50 ms) with a blank period of one monitor frame (16.67 ms) in between items.

The cue appeared around a target letter in one of the streams, chosen randomly on each trial. The cue was presented for the same three monitor frames as the target letter. The target letter was at a random serial position in the stream between the seventh and the tenth items, inclusive. Streams were equally spaced about an imaginary circle with a radius of 3.0º, centered on fixation.

In the eight-stream condition, the equal positioning of stimuli about the circle resulted in a centre-to-centre spacing of 2.3º of visual angle. In the two-stream condition, the streams were presented on opposite sides of fixation with a separation of 6º of visual angle. With that constraint, across all experimental trials the two streams occupied all eight positions occupied by the eight-streams stimuli. The two streams appeared in every position the same number of trials across the experiment.

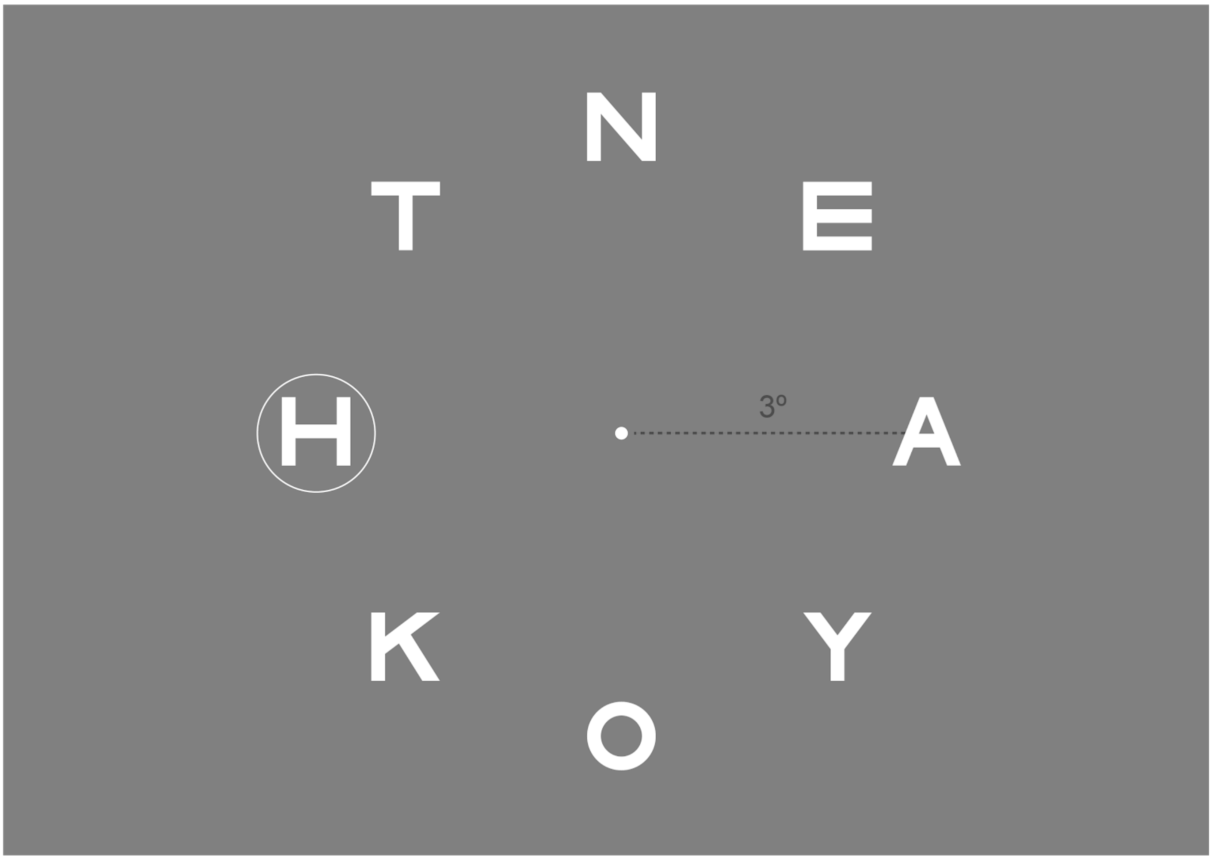

Figure 5. A schematic display of an eight-streams display, showing the cue.

At the end of each trial, a response array appeared that contained all 24 letters. Participants indicated the target by selecting a letter with the mouse. They were told to click with the right mouse button if they were “sure” and the left mouse button if they were “unsure” of their response.

Procedure. Participants completed 320 trials. Half were the two-stream condition and half were the eight-stream condition, randomly intermixed. On-screen instructions informed participants that on each trial they must keep their eyes fixed on the central point while “several rapid, randomly-ordered sequences of letters” appeared at two or eight locations on the screen. The instructions also informed participants that one letter would appear “with a white ring around it” and that their task was to report this letter by clicking on it in the response array.

Analysis. The first twenty trials were treated as practice and excluded from the analysis. Mixture models were fit to the remaining 300 trials with a custom R package (https://doi.org/10.5281/zenodo.3545085). To compare parameters between the two and eight streams condition, we computed Bayesian t-tests with JSZ priors (Rouder et al., 2009) using the BayesFactor package (Morey & Rouder, 2018) in R (R Core Team, 2019).

To assess how much the data of each participant, and in each condition, favored the Gaussian or skewed gamma distribution, we estimated the Bayes factor – the probability of the data according to one model divided by the probability according to the other model. This Bayes factor, with an uninformative prior over the model parameters, can be estimated using the Bayesian information criterion for each model (Raftery, 1999; Wagenmakers, 2007). We calculate the BIC for each model and the Bayes factor estimated from the BICs to assess which model is a better fit to each participant’s data. This mixture model comparison gives us a ratio for the evidence of one model relative to the other. Bayes factors with a ratio greater than three in favour of a particular model are taken as evidence for that model.

4.5.2 Results

Fixation data were retrieved from the Eyelink 1000 using the fixation report from Eyelink Data Viewer (Version 1.11.900)**.** Trials on which the participant made a fixation more than 1° of visual angle from the centre of the fixation point were excluded from further analysis. A mean of 16.9 (SD = 10) trials per participant were rejected in the two-streams condition. A similar mean of 14.9 (SD = 7.4) trials were rejected in the eight-streams condition.

Sure/unsure responses. The statistics reported below indicate that when participants made a “sure” rather than an “unsure” response, they were more likely to have reported the target or a letter close in time to the target. There was, however, substantial variance in the number of “sure” responses, and some participants responded “sure” only rarely. The number of “sure” responses ranged from 1 to 61 with eight streams and 1 to 64 with two streams, corresponding to 0.007 to 0.46 and 0.007 to 0.45 of the trials in which participants maintained fixation in each condition, respectively. We fit a set of mixed effects logistic regressions predicting the odds of responding “unsure” with condition and the absolute value of SPE as predictors (see Appendix). The absolute value of SPE was chosen as an independent variable to assess whether participants have insight into the extent to which their response differs from the target. These data were best explained by a model with two predictors, the absolute value of the SPE on each trial and the condition. The odds of responding “Unsure” increased by 25% for every serial position increase in distance from the target item (b = .23, z= 8.008, p =1.17 x 10^{-15}). A similar result of confidence decreasing with the absolute value of SPE was reported by Botella (1992). The odds of an unsure response in the eight-streams condition were 64% of the odds in the two-streams condition (b = -.45, z = 4.57, p = 4.8 x 10^{-6}). That is, participants were *less* likely to make an “unsure” response with eight streams than with two.

The number of “sure” responses were too few to reliably estimate the parameters of the mixture model for the different confidence responses, so “sure” and “unsure” responses are collapsed together for the mixture model and binomial analyses.

Model Comparison. The data strongly favored the Gaussian symmetric model with two streams, and only with eight streams did they favor the skewed gamma model (Figure 6). The Bayes factors indicated strong evidence for the Gaussian model for nine out of ten participants in the two-streams condition – Bayes factors ranged from 378 to 1.05 x 10^{15} in favour of the Gaussian model. One participant's data did not favour either model (1.02).

Evidence for skew (the gamma model) was seen only in the eight streams condition, where data from three participants was fit best by the gamma model (BFs: .05, .26 and .17). The remaining seven participants had model fits that did not markedly favor either model (BFs ranging from .38 to 2.95).

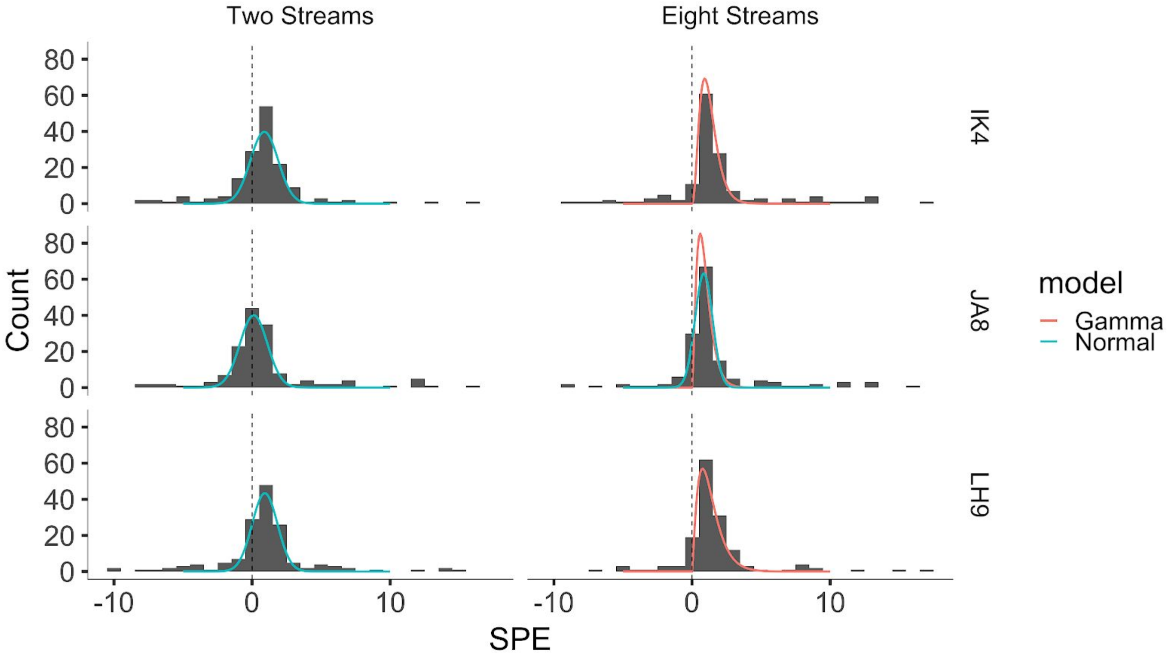

Figure 6. Serial position error (SPE) histograms and model fits for three randomly-selected participants (rows) in the different conditions (columns) in Experiment 2. The density of the best fitting efficacious distributions are shown as the lines on each plot, scaled to near the height of the histograms. For one of these participants, the data did not allow us to discriminate between the Gaussian and gamma models, and the density curves from both efficacious distributions are plotted. For two of these participants, the skewed gamma distribution fit significantly better than the symmetric Gaussian distribution in the eight stream condition.

Parameter Estimates. Because for the majority (seven of ten) of the participants, the data for the eight streams condition did not strongly favor either model, we follow Goodbourn and Holcombe (2015) and interpret the parameters of the Gaussian mixture model for both conditions. Fortunately, the pattern of the parameter values (described below) is the same when the parameters of the skewed (Gamma) model are used instead. The exception to this is the gamma mixture's efficacy estimates, which are higher for the eight stream condition than for the two streams condition. This is likely to be because with few streams, participants made efficacious selections of items from before the cue, which are not captured by the gamma distribution (it has no mass before zero).

Efficacy. Efficacy is the proportion of responses that are efficacious. The mean estimated efficacy for the two-streams condition is 0.74 (SD = 0.13), and the mean for the eight-streams condition is a very similar 0.76 (SD = 0.11), yielding tentative evidence for no difference, according to the Bayesian paired t-test (BF_{10} = 0.38, *d* = 0.15, 95% CI [-0.33, .65]).

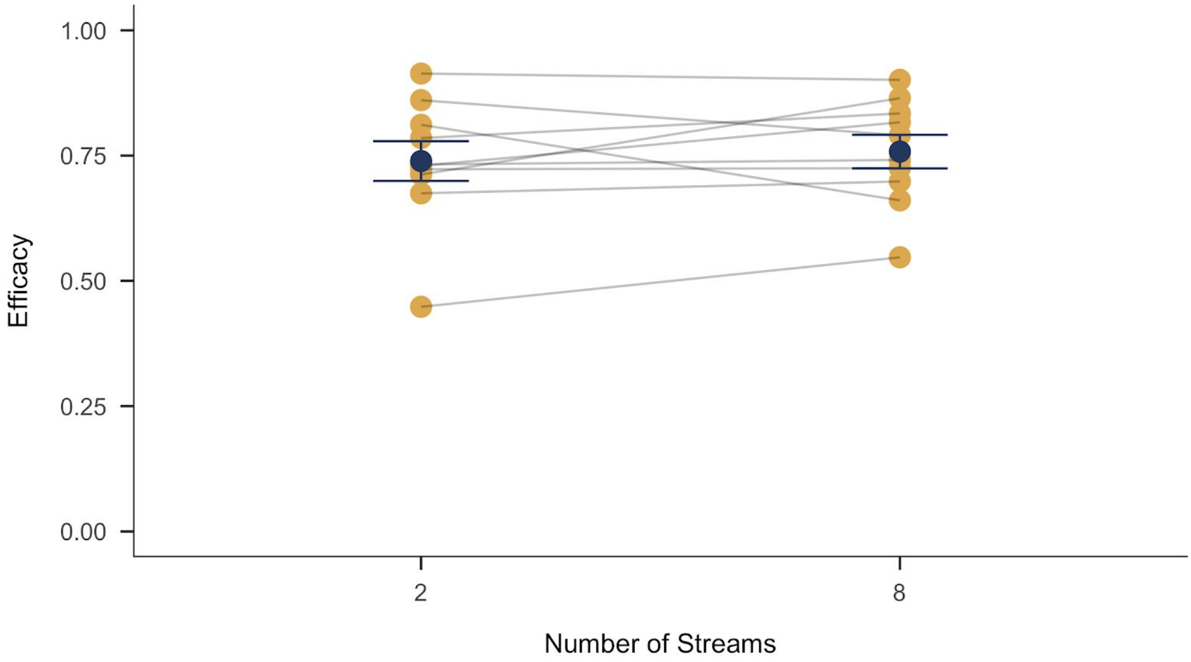

Figure 7. Efficacy estimates for the two and eight streams conditions. The dark points and error bars indicate the mean ±SE for each condition. The lighter points and lines represent each participant's estimate.

Latency. The latency of the best-fitting efficacious distribution is longer in the eight streams (M = 72.5 ms, SD = 16.6) than in the two streams condition (M = 40.0 ms, SD = 18) according to a Bayesian paired t-test (BF_{10} = 241, *d* = 1.87, 95% CI [0.86, 2.89]).

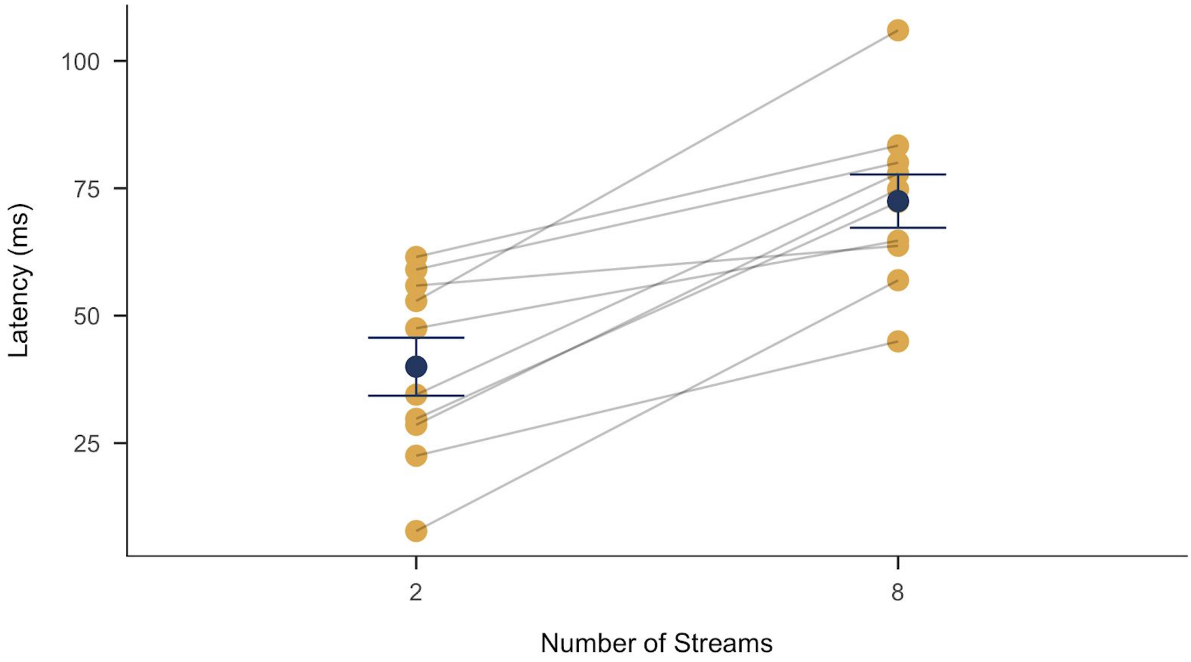

Figure 8. Latency estimates, the mean time of the efficacious distribution, for the two and eight streams conditions. The dark points and error bars indicate the mean ±SE for each condition. The lighter points and lines represent each participant's estimate.

Precision. Precision was smaller (the efficacious distribution was narrower) for the eight streams conditions (M = 41.3, SD = 10.5) than for the two -streams condition (M = 61.2ms, SD = 13.3), Bayesian paired t-test BF_{10} = 45.9 (*d* = 1.64, 95% CI [0.56 2.72]).

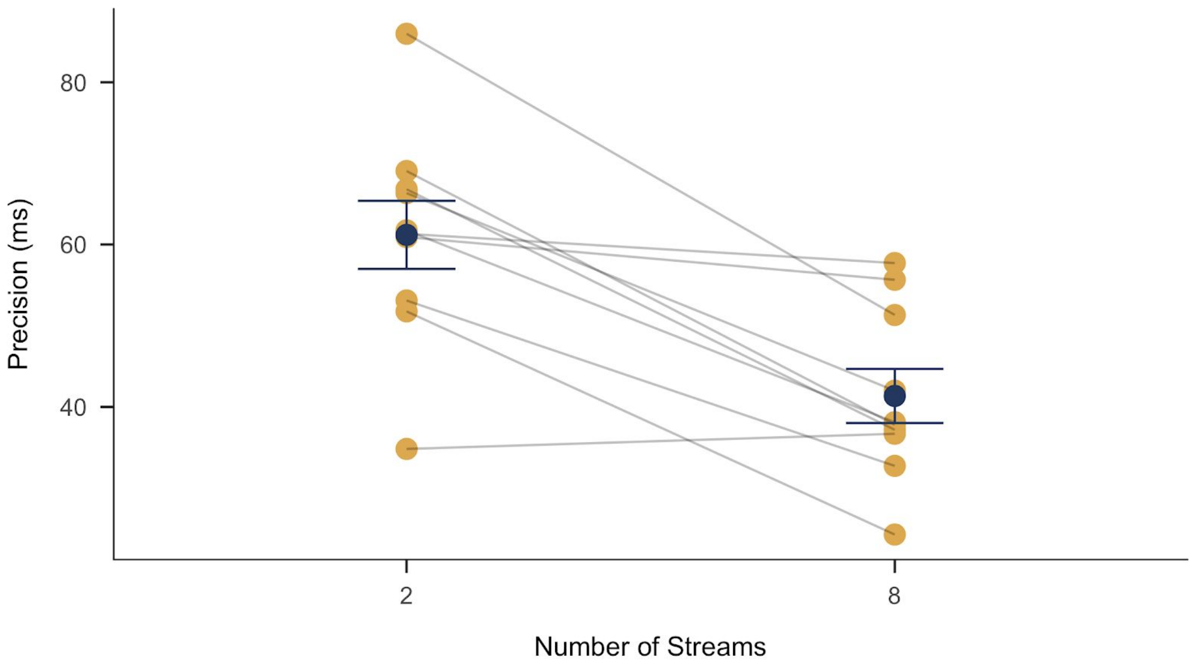

Figure 9. Precision (the standard deviation of the efficacious distribution) for the two and eight streams conditions. The dark points and error bars indicate the mean ±SE for each condition. The lighter points and lines represent each participant's estimate.

Binomial Test. The binomial analysis indicates that in the two streams condition, eight of the ten participants had more SPEs of -1 than expected from our liberal estimate of the rate of identification failures. *p*-values for these participants ranged from 1.9×10^{-11} to 0.03. In contrast, in the eight streams condition, the null hypothesis that the SPEs of -1 are explainable by identification failures was not rejected for any of the participants.

4.5.3 Discussion

The experiment yielded evidence for buffering in the two stream condition but not the eight stream condition. In addition, in the eight stream condition, the temporal distribution of efficacious reports occurs later, is narrower, and is less likely to include reports of items from before the time of the cue.

Participants were more confident in their responses when there were eight streams than when there were two. This finding is similar to observations that participants tend to be more confident in their responses for difficult perceptual tasks relative to easier tasks (Baranski & Petrusic, 1994). Increasing the set size in a visual search task - a manipulation similar to our increase in the number of streams - results in more confident responses (Baldassi, Megna, & Burr, 2006). This effect is typically thought to reflect decision processes or a lack of insight into task difficulty (Suantak, Bolger, & Ferrell, 1996).

The apparent absence of buffered reports (efficacious reports of items from before the cue) in the eight streams condition is consistent with our hypothesis that buffered responses would be less frequent as the number of streams increased. Moreover, the evidence sometimes favored a skewed efficacious distribution in the eight stream condition, which is consistent with selection occurring via attention shifts.

A finding we were surprised by is that participants appear just as effective (equal efficacy) at selecting an item around the time of the cue in the eight stream condition as in the two stream condition. We thought that when several streams were presented, the selection process would shift from often accessing information from a buffer to requiring an attention shift. The evidence for no change in efficacy is surprising under this theory, because the two processes should have different efficacies. However, we now believe that buffered responses result from changes in the speed of attention’s engagement following the cue. When there are few streams, each stream receives more attentional resources and attention is faster to select an item in the cued stream than when there are many. Efficacious pre-cue reports occur in these conditions because the speed of attention is such that it may select a stimulus representation that has not yet decayed despite masking. This explains the lack of a change in efficacy, because in both the two- and eight-streams conditions selection occurs via the same mechanism - an attention shift to the cued location.

In Experiment 2, we further assessed the extent to which efficacy is resilient to an increasing number of streams by increasing the number of simultaneous RSVP streams to 18.

4.6 Experiment 2

4.6.1 Method

Participants. Thirteen participants took part in the experiment. Nine were undergraduates (six female) aged 18-21 years (*M* = 19.9 years, *SD* = 1.4) and four were graduate students (two female). The age of the graduate students was between 25 and 32. All were students at the University of Sydney. Eye movements for one participant were not recorded due to a computer error, so this participant was excluded from the analysis. This experiment was not pre-registered. The same Bayesian stopping rule as Experiment 1 on the latency parameter was used. Data collection stopped when the ratio of evidence for an effect of the number of streams to the null hypothesis exceeded ten in favour of either hypothesis.

Stimuli. Participants viewed two, six, or 18 streams on each trial. The 18 streams condition comprised three concentric rings of six equally-spaced streams. The eccentricities of the rings were 3, 7 and 11.5 degrees. The ring with radius 7° was rotated clockwise by half the polar angular separation between the streams. The six streams condition consisted of six stimuli with equal eccentricity spaced around a circle. In the two streams condition, the streams were presented on opposite sides of fixation, at an angle randomly chosen from those occupied the streams of the 18 streams condition. Each stream consisted of the same 24 letters used in Experiment 1 with no repeats, presented in a random order.

Stimuli were eccentricity scaled with the parameters used by Strasburger (2005) for scaling numeral stimuli, resulting in letter heights of 0.9°, 1.62° and 2.43° for stimuli at 3, 7 and 11.5 degrees of eccentricity respectively. Cue diameter was scaled from 0.98º at 3º of eccentricity to 1.77º and 2.65º at 7º and 11º, respectively. The number of presentations of the cue at a particular location on the screen (and thus a particular eccentricity) was equal across conditions.

Stimuli were presented at a rate of approximately 12.5Hz – each stimulus was presented for 6 monitor frames (60 ms) with a blank period of 2 monitor frames (20 ms) between each item. The target letter was a random item in the stream between the 7th and the tenth items, inclusive.

Apparatus. Participants viewed stimuli in a dark room on a Mitsubishi Diamond Pro monitor from a distance of 42 cm, ensured by a headrest. The monitor refreshed at a rate of 100Hz. Movements of the right eye were measured with an SR Research Eyelink 1000.

Procedure. Thirteen undergraduate participants performed 270 trials, 90 for each number of streams (2, 6, and 18), randomly intermixed. One participant was dropped from the analysis because an error resulted in their session not being eyetracked.

Analysis. Mixture model fits and model comparisons were computed in the same manner as Experiment 1. Parameter estimates were compared using Bayesian within-subjects ANOVA implemented in the BayesFactor Package (Morey & Rouder, 2018) in R (R Core Team, 2019)

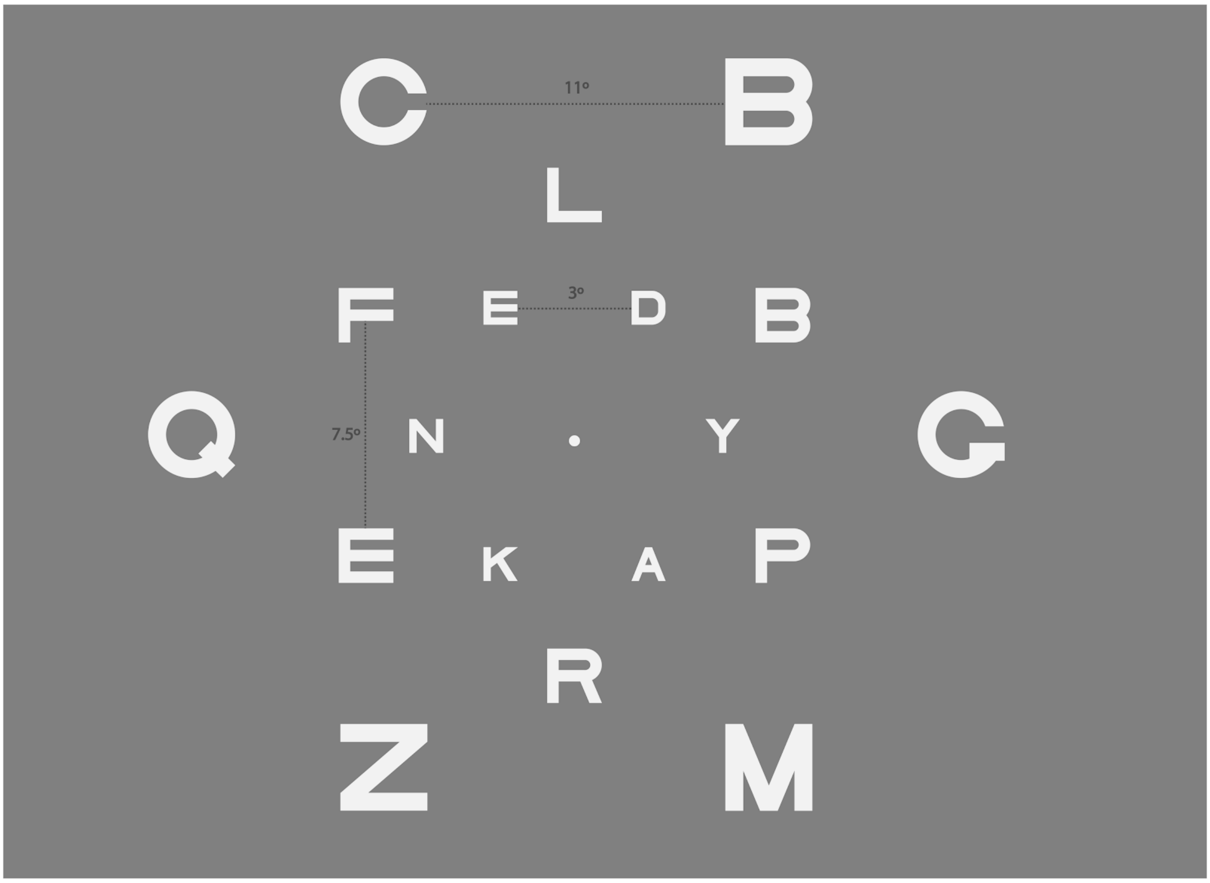

Figure 10. A schematic of a display from the 18 streams condition. Stimuli were arranged into three rings of six, equally-spaced streams each. Measurements in degrees represent the centre-to-centre spacing of objects in the same ring

4.6.2 Results

The mean number of trials discarded because of the presence of a fixation that fell more than 1° from the fixation point were 13.9 (SD = 11.4), 12.4 (SD = 10.0) and 15.3 (SD = 9.2) for the 2, 6 and 18 streams conditions, respectively.

Model Comparisons. Participants' data in the two-streams condition all favoured the Gaussian mixture model over the skewed Gamma model with the exception of one participant (Table 3). The Gaussian model was also favoured in the six streams condition by the data of those participants whose data had Bayes factors that were not ambiguous. In the eighteen streams condition, evidence for the skewed gamma distribution emerged, with the data of five of the 12 participants. Example histograms and efficacious distributions are presented in Figure 11.

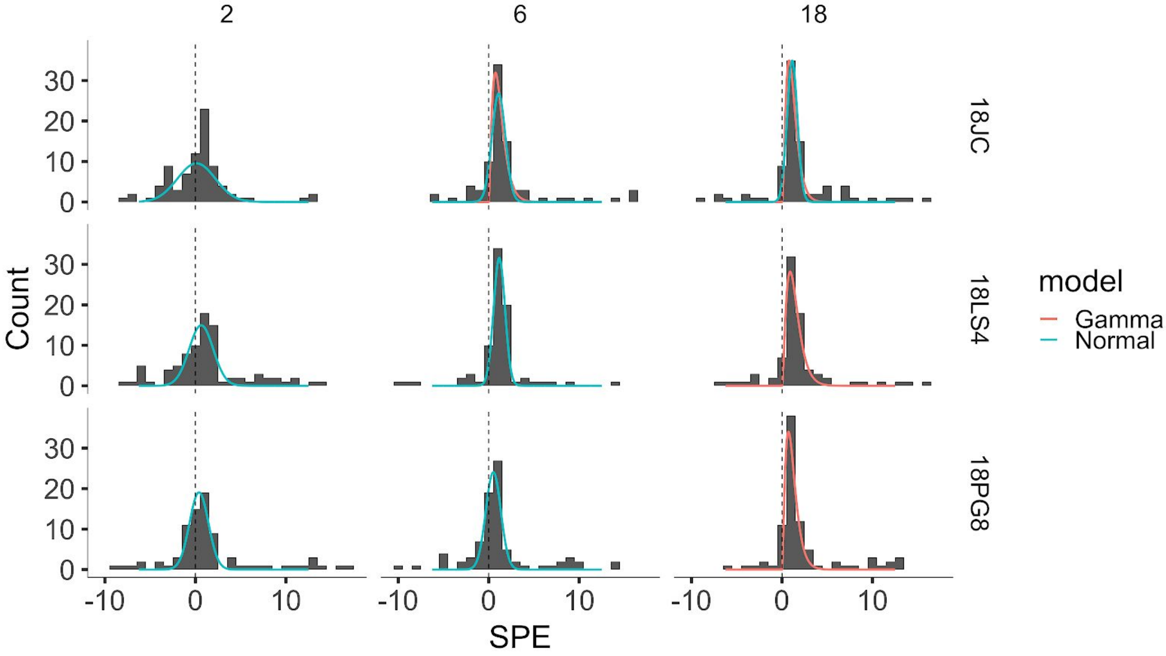

Figure 11. Serial position error (SPE) histograms and models fits for three randomly-selected participants (rows) from Experiment 2 for different numbers of streams (columns). The density of the best fitting efficacious distributions are shown with colored lines, scaled to near the height of the histograms. For participants whose data did not allow us to discriminate between the buffering and attention shift models, density curves from both efficacious distributions are plotted. For two of these participants, selection from one of 18 streams yields data that strongly favors the skewed (attention shift) distribution, so only that density is shown.

Table 3

The number of participants whose data favoured the symmetric (Gaussian) model versus the skewed gamma model.

Condition	Gaussian	Gamma	Neither
Two Streams	11	0	1
Six Streams	6	0	6
Eighteen Streams	2	5	5

Parameter Estimates. The parameter estimates are presented in Table 2.

Table 2

Means and standard deviations for the parameter estimates for each condition

Number of Streams	Efficacy (SD)	Latency (SD)	Precision (SD)
2	.73 (.11)	24.1 (31.9) ms	111.0 (50.0) ms
6	.74 (.09)	67.6 (30.9) ms	62.5 (17.5) ms
18	.74 (.08)	90.9 (27.1) ms	56.7 (9.74) ms

Efficacy. A Bayesian repeated measures ANOVA indicates no difference in efficacy across the three conditions (BF_{10} = .19, η^2_G = 0.003)

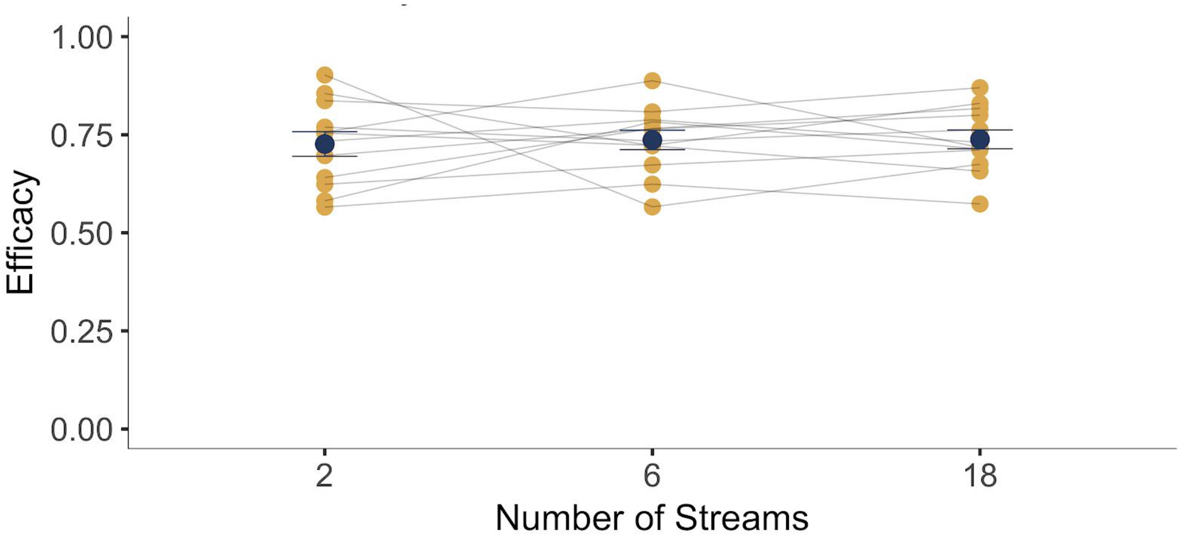

Figure 12. Efficacy estimates for the different conditions in Experiment 2. Efficacy was not affected by the number of streams.

Latency. Latency increased with the number of streams (η^2_G = 0.46). A Bayesian test of the ordering of the latency estimates in the same direction as Experiment 1 (2-streams <

6-streams < 18-streams) conducted in the manner described by http://bayesfactor.blogspot.com/2015/01/multiple-comparisons-with-bayesfactor-2.html found that this ordering was 7.8×10^6 times better able to explain the data than the null model and 5.98 times more likely than the model with no order restriction.

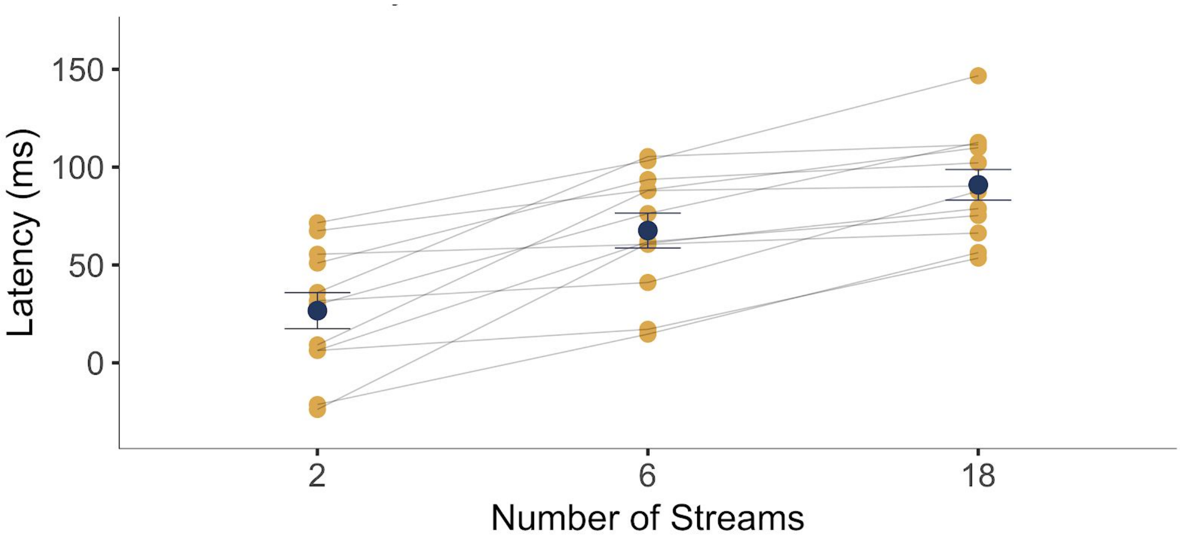

Figure 13. Latency estimates for the different conditions of Experiment 2. Latency increases with the number of streams.

Precision. The distributions narrowed (the precision estimates decreased) as the number of streams increased ($\eta^2_G = 0.40$). The data were most likely under a model with an order of precision estimates in the same direction as experiment 1a (2-streams > 6 streams > 18 streams). The data under this model were 1905 times more likely than under the null model and 4.1 times more likely than under a model with no order restriction.

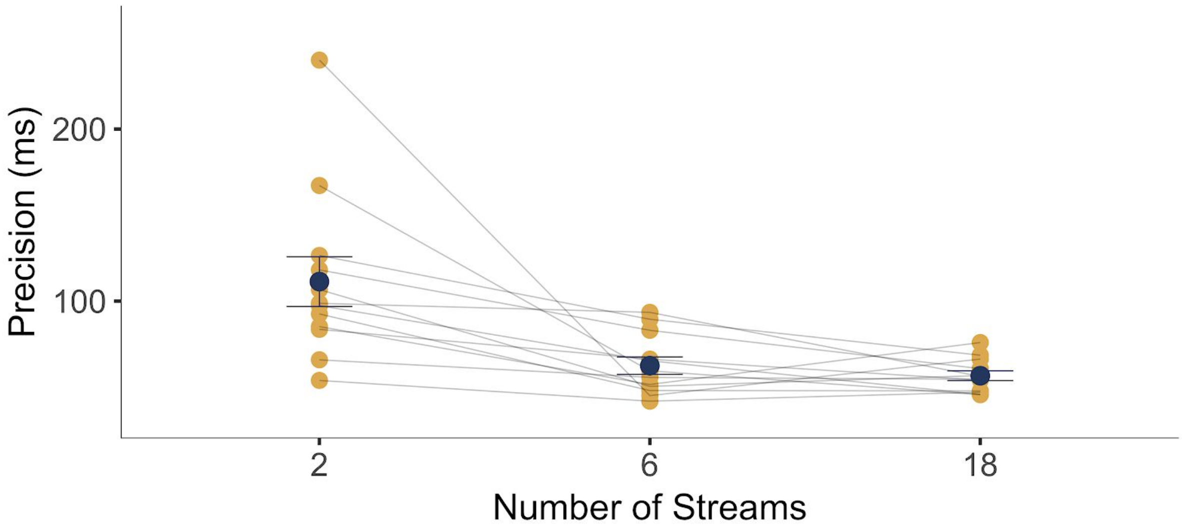

Figure 14. Precision estimates for the different conditions in experiment 2. Precision narrows (the estimates decrease) as the number of streams increases.

Binomial Test. Ten of the 12 participants made significantly more SPE = -1 responses than predicted by our liberal estimate of identification failures in the two streams condition ($2.5 \times 10^{-11} \leq p \leq 0.02$), four in the six streams condition ($1.2 \times 10^{-5} \leq p \leq 0.01$) and none in the 18 streams condition (all *p*s > 0.3).

4.6.3 Discussion

The pattern of results is consistent with those of Experiment 1. The binomial test and model fits found evidence for buffering only with few streams. Efficacious selections were delayed and less variable in time as the number of streams increased, but there was no change in the efficacy of selection.

The lack of change in efficacy argues against separate processes for selection from the buffer and attentional selection from the cued stream proposed by Goodbourn and Holcombe (2015). This pattern of results, observed in the current experiment and Experiment 1, suggested to us that efficacious pre- and post-cue reports are the result of the same mechanism. One possible explanation is that participants are dedicating attention to the locations of the streams before the cue occurs. Attending to few streams, relative to many, at a location may have a number of benefits when it comes to detecting and responding to the cue (Castiello & Umiltà, 1990, 1992). With few streams, there are more attentional resources are endogenously allocated to each stream prior to the onset of the cue. This may result in selection in response to the cue occuring more quickly than when attention is diluted among many streams. Under this account, participants are less likely to make an efficacious response of a pre-cue letter as the number of streams increases because their attentional resources are spread over a larger area, causing cue-triggered selection to be delayed.

Both pre- and post-cue reports seem to be selected with the same process. Efficacy is unchanged with few streams where there are efficacious pre-cue reports relative to when there are many and the efficacious distribution is delayed. In conditions where attention engages quickly with the stream, it appears that it can sometimes select the representation of a pre-target stimulus if that stimulus' representation has not yet decayed. Thus, despite the masking inherent in RSVP, letter representations persist beyond the presentation of a subsequent letter. This result is inconsistent with assumptions about the effect of marking in certain published theories of RSVP. Some models assume that stimulus representations are terminated when a new stimulus is presented at the same location (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002).

4.7 Experiment 3

In the preceding discussion, we attributed the effects of increasing the number of streams as due to there being more streams in which the target might occur. But the effects may instead have occurred simply because there were more streams physically present on the screen. The presence of more streams might result in changes to the dynamics of selection through interference among stimulus representations, even if those streams were not potential locations for the cue and target.

In some circumstances, simultaneously presented visual stimuli certainly do interfere with each other. One form of interference is known as “crowding”. In crowding, flanking stimuli markedly impair identification performance (Agaoglu & Chung, 2016; Bouma, 1970; Pelli & Tillman, 2008; Rosenholtz, Yu, & Keshvari, 2019; Whitney & Levi, 2011). The size of the region in which stimuli crowd each other scales approximately linearly with eccentricity. Stimuli with a centre-to-centre spacing of less than 0.4 to 0.5 of their eccentricity will be crowded (Pelli & Tillman, 2008). However, it is unlikely that crowding can explain our results because our stimuli were widely spaced enough to fall outside of the crowding zone. The centre-to-centre spacing in the eight-streams condition of Experiment 1 is 0.77 of the eccentricity. The closest items in the 18-streams display of Experiment 2 are those in the inner ring, which have a spacing of 3°, the same as their eccentricity.

Although crowding is unlikely to have occurred in our experiments, even stimuli with large spacings may compete for processing resources that are limited (Desimone & Duncan, 1995). To investigate whether this is the reason for the effect of number of streams in Experiments 1 and 2, we hold constant the level of potential interference between stimulus representations in Experiment 3 by always presenting the same number of streams on each trial. We pre-cued the potential locations of the target on each trial and varied the number of potential locations. Any effect of number of potential locations can be attributed to dilution of attentional resources, with less resources devoted to each stream.

Eight streams were presented on each trial, and participants were instructed to monitor two or eight streams for the cue. This mimics experiment 1 but controls for any interference effects as there are always eight streams. Thus any differences we observe in the latency and precision of these results are due to changes in attentional spreading rather than interference, which is fixed on these trials. A replication of the pattern of findings in experiment 1 would be evidence for the idea that these effects reflect participants spreading endogenous attention over many locations for the presence of the cue, rather than an effect of the amount of interference between stimuli.

4.7.1 Method

The method, analysis and apparatus (including eyetracking) for this experiment were the same as that of Experiment 1, including the Bayesian stopping rule (10:1 in favour of the null or the alternative hypothesis), with the exception that eight streams were always displayed and a precue indicated the possible spatial positions of the cue on each trial. Prior to the presentation of the RSVP streams on each trial, eight hash marks appeared on the screen for 250ms, occupying the same positions as the RSVP streams. Two or eight of these hash marks had white rings around them, representing the possible positions of the cue on that trial. The two-precue condition precued the same locations as those occupied by the two streams in Experiment 1 (either side of fixation). A blank screen was presented for 500 ms after the precues. 13 undergraduate students from the University of Sydney participated and received course credit. This experiment was not preregistered.

4.7.2. Results

A mean of 18 (SD = 13) trials were rejected based on eye movement away from fixation in the two-precue condition. A mean of 15 (SD = 12) trials were rejected in the eight-precue condition.

Model Comparisons. Model fits largely favoured the Gaussian mixture model or did not distinguish between the models. See table 4 for the frequency of each fit in each condition.

Table 4

The results of the model comparisons. The number of participants whose data favoured each model.

Condition	Gaussian	Gamma	Neither
Two Precues	6	1	6
Eight Precues	6	2	5

Parameter Estimates.

Efficacy. The mean efficacy in the two-cues condition was 0.75 (SD = 0.07), and in the eight-precues condition was 0.71 (SD = 0.13; BF_{10} = .65, *d* = 0.28, 95% CI [-0.13, 0.70]). A Bayesian paired t-test comparing efficacy estimates between conditions produced a Bayes factor close to 1, indicating that the test favored neither the null hypothesis of no difference nor the alternative hypothesis of a difference.

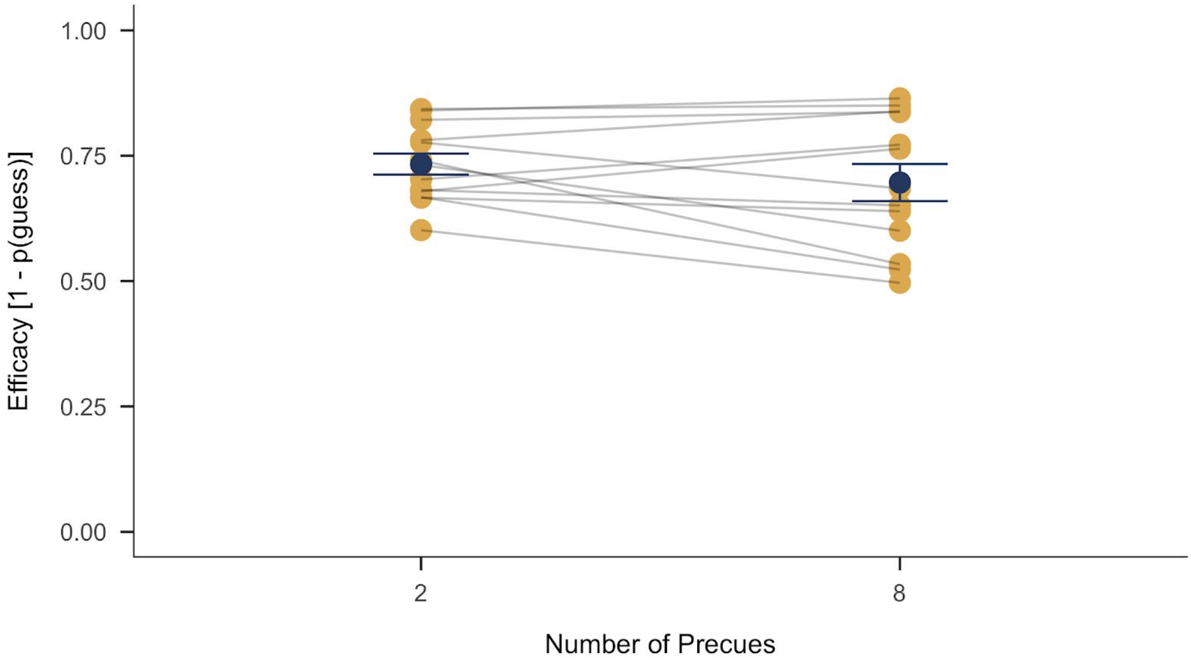

Figure 15. Efficacy estimates for Experiment 3

Latency. Latency was shorter for the two-precues condition (M = 84.84, SD = 72.35) than the eight-precues condition (M = 108.12, SD = 69.37; BF_{10}= 21.9, *d* = 0.33, 95% CI [0.15, 0.50]). One participant had very high latencies - around 280 ms - in both conditions and also large precision estimates of 180 ms and 162 ms in the two and eight-precue conditions, respectively. But removing this participant from the efficacy, latency and precision analyses did not change the pattern of results.

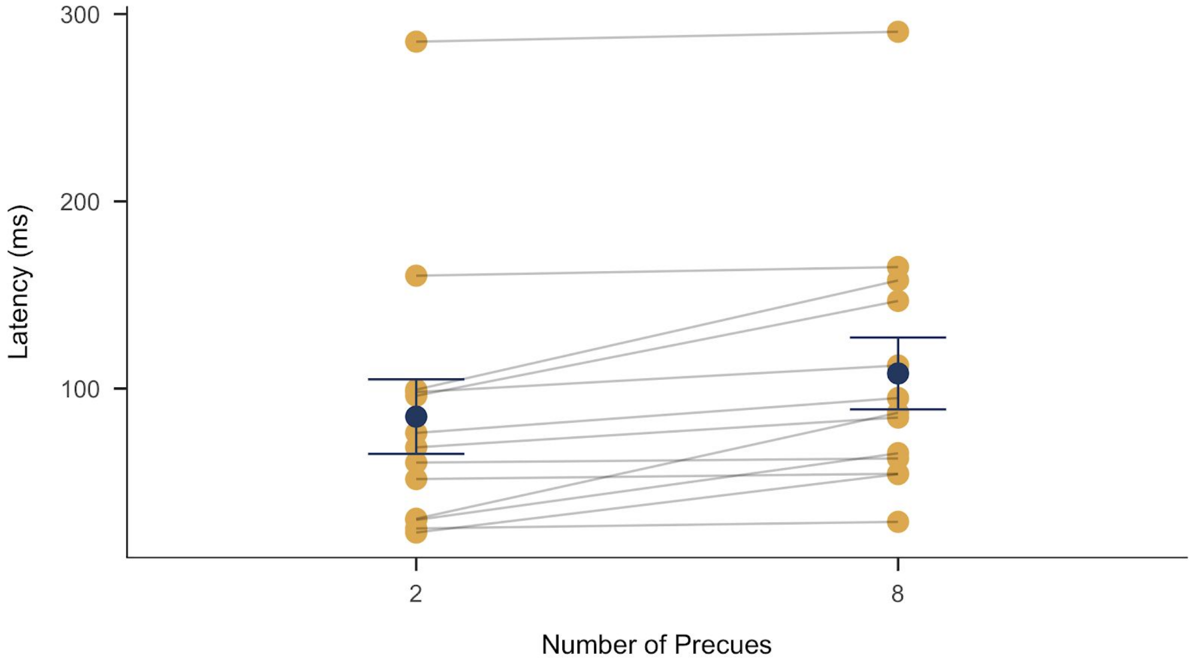

Figure 16. Latency estimates for Experiment 3

Precision. The null hypothesis of no no difference in precision between the two conditions explained the data best, BF_{10} = 0.35. The two-precues condition had a mean of 90.3 ms (SD = 49.5). The eight-precues condition had a mean of 85.5 ms (SD = 39.3, *d* = 0.08, 95% CI [-0.16, 0.33]).

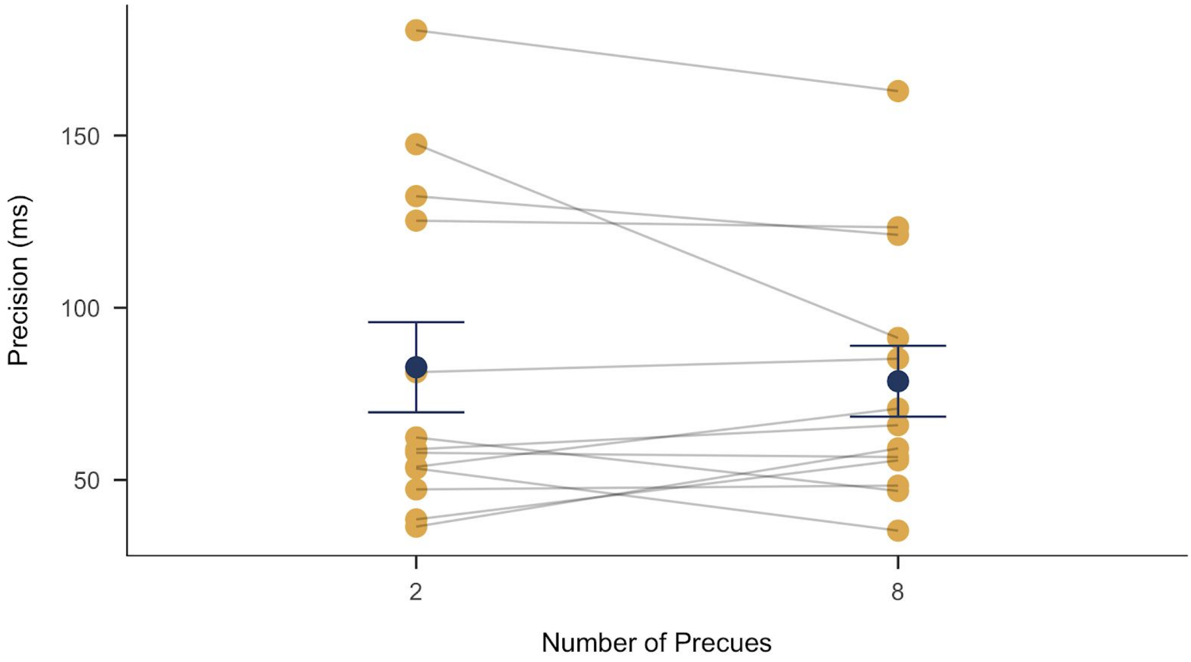

Figure 17. Precision estimates for Experiment 3

Binomial Test. Three participants in the two-precues condition had more SPEs of -1 than are easily explained by identification failures (*p*s < .001). One participant showed more SPEs of -1 than would be predicted by identification failures in the eight-precues condition (*p* = .01).

Comparison of effects between experiments (Exploratory Analysis). To compare the effect of pre-cueing and the effect of the number of streams, we combined the data from Experiments 1 and 3 and tested for differences between them in an exploratory analysis. We fit two linear mixed-effects models for each of the three mixture model parameters using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). The models always contained random intercepts by participant and predicted each parameter with main effects of condition and experiment, or the main effects and an interaction between condition and experiment. This interaction term represents a test of the hypothesis that changing the number of pre-cued streams (Experiment 3) is different from changing the number of streams (Experiment 1). If that is the case for a given parameter estimate, then the best fitting model for that estimate should contain an interaction term. To assess model fit, we calculated Bayes factors from each model's BIC using the approximation described by Wagenmakers (2007).

The Bayes factors produced by this analysis represent the ratio of evidence for the interactive model to the model without an interaction. This analysis indicated no change between experiments in the effect of condition on efficacy (BF_{10} = .46) but was inconclusive for latency (BF_{10} = 1.13). For precision, there is some evidence for a difference between the experiments – more streams narrowed the efficacious distribution in Experiment 1 but not Experiment 3 (BF_{10} = 3.77).

The effect of spreading attention over many streams, as manipulated by precuing, does not completely explain the changes in temporal processing in experiment 1 because the effect of condition on precision differs between the experiments.

4.7.3 Discussion

By manipulating the number of relevant streams, while keeping the number presented constant, we replicated the findings of Experiment 1 in that more candidate streams yielded longer latencies and fewer efficacious reports of the item before the cue. Our test that Experiment 3's latency change was quantitatively the same as that observed in Experiment 1 yielded equivocal evidence. However, both experiments produced strong evidence for an increase in latency as either the number of streams increased (Experiment 1) or the number of relevant streams increased (Experiment 3). This suggests that the increase in latency in Experiment 1 is due, at least in part, to changes in the number of locations that participants must dedicate their attention to. When participants spread their attention over many simultaneous streams, they are slower to orient their attention to the cued location than when there are fewer streams. Not all the changes in the temporal qualities of selection can be explained by participants spreading attention over the potential locations of the cue, however. There was no change in precision for different numbers of streams in Experiment 3, whereas in Experiment 1 precision was narrower for higher numbers of stream. The evidence for this difference between the experiments was only moderate, however, so more work is needed to confirm it.

The proportion of participants for which there was evidence of efficacious responses of pre-cue items was higher when there were only two pre-cued locations, suggesting that such responses reflect the application of endogenous attention over the possible locations of the cue. In other words, reports of letters from before the cue that are efficacious are more likely to occur when participants monitor a small number of locations for the cue.

4.8 General Discussion

The present experiments and analyses provide the strongest evidence to date for efficacious responses of pre-cue items despite the presence of masking in a cued RSVP task. Our binomial procedure tests whether pre-cue responses are more frequent than predicted by identification failures. Our mixture models estimate the rate of identification failures and allow us to assess changes in the latency and shape of the selection distribution across conditions.

Together these analyses indicate that participants often make an efficacious response of a pre-cue item when there are few streams. This is consistent with previous suggestions that such responses may occur (Botella et al., 2001; Goodbourn & Holcombe, 2015; Holcombe et al., 2017; Vul & Rich, 2010), but is the first to directly test whether the number of responses from before the cue could be explained by identification failures.

By investigating this issue with more than two RSVP streams, we discovered that evidence for efficacious pre-cue responses diminishes rapidly with more streams. Previous theories of human performance in RSVP either cannot account for this finding, in the case of theories of the attentional blink and theories that assume some sort of buffer, or are explicitly falsified if they assume that stimulus representations are overwritten due to masking (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002). We suspect that when only a few streams are presented, attention is engaged with those streams prior to the cue, so that attention swiftly orients to the target's location. We argue below that in this situation, the item before the cue can be reported because the speed of selection is sometimes fast enough to select a pre-cue item's representation before it decays.

Experiments 1 and 2 demonstrated pre-cue responses only with few streams, but using larger numbers of streams in those experiments potentially also altered the interference among stimuli in the display. To avoid this confound, we held the number of streams presented constant while changing the number of potential cue locations (Experiment 3). In this experiment the number of responses from before the cue dropped as the number of potential cue locations increased. This suggests that pre-cue responses are the result of the voluntary application of attention to the potential cue locations. Participants can control the application of attention to the relevant streams, and this affects the speed of selection as well as the likelihood of reporting an item from before the cue.

Goodbourn and Holcombe (2015) assumed that buffering was distinct from the process of attention shifts to the location of the cue, based on their (untested) observation of efficacious responses from before the cue. Our results, however, argue against separate buffering and attentional selection processes despite the presence of pre-cue responses. Selection may be parsimoniously explained as the result of a single process in all cases: exogenous attentional selection triggered by the cue. It does not appear that efficacious pre- and post-cue responses are produced by separate processes because there are no systematic changes in efficacy as the number of streams changes in our experiments. If a different process were involved when there were few streams, we would expect to observe a change in efficacy between conditions because different processes seem unlikely to have the same probability of success. There is no such change here, and thus items are selected by attention in all conditions.

We suggest that selection in all cases here is a result of cued attentional shifts. Increases in the selection latencies occur because participants apply endogenous attentional resources to the RSVP streams prior to the onset of the cue. Thus, the time of selection is a function of the attentional resources endogenously dedicated to the cued stream prior to the cue. When there are few streams, each stream receives a larger proportion of the available attentional resources. This speeds selection at the cued location relative to conditions with greater numbers of streams. That this pre-cue resource allocation is endogenous is supported by our finding that latency is manipulated by providing participants with information about the upcoming location of the cue (Experiment 3).

One may expect that masking would terminate the representation of the SPE -1 stimulus, eliminating the possibility that attention selects that item. Several models of selection in RSVP tasks assume this (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002) based on observations that attentional selection of from sensory memory is not observed when stimuli are followed by a postmask (Averbach & Coriell, 1961; Sperling, 1960). However our results indicate that sensory information about stimuli persists in RSVP despite masking and can be selected by attention, falsifying this assumption. This claim is supported by recent neural decoding work demonstrating that information about RSVP stimuli persists beyond the presentation of a subsequent stimulus at the same location (King & Wyart, 2019; Marti & Dehaene, 2017) and behavioural work indicating that selection from a post-masked static display is possible

(Smithson & Mollon, 2006). However, our results are the first to show that such information can inform behaviour in RSVP.

Our theory explains the pattern of results found for latency, efficacy, and precision. As explained above, latency increases because with more streams, the endogenous application of attention must be spread over more locations, leading to a smaller attentional advantage for each stream and delaying selection. Diluting attentional resources changes selection latency, but not the probability of detecting the cue and identifying a letter. The latter process always proceeds via the same mechanism, and thus efficacy does not change across conditions in the three experiments.

The temporal distributions of the serial positions reported were narrower with larger numbers of streams in Experiments 1 and 2, but not in Experiment 3 where the number of streams was constant but the number relevant to the task was varied. Perhaps there is competition among streams for cortical processing, whether they are relevant or not, and this causes stimulus representations to not persist as long as with fewer streams. The range of presentation times of stimuli that are simultaneously activated, and available for selection, should then be narrower when more streams are presented. One caveat is that the previous evidence for long-range interference in visual cortex is thin. The relevant neuroimaging studies that we are aware of all spaced stimuli closely enough that stimuli may have crowded each other, unlike our RSVP streams (Beck & Kastner, 2005, 2007; Kastner et al., 1998, 2001; Scalf & Beck, 2010). However, there is some neurophysiological evidence for long-range interference (Falkner, Krishna, & Goldberg, 2010; Schall et al., 2004; but see Holcombe, Chen, & Howe, 2014 on interpretation of these).

This explanation for the change in precision may seem inconsistent with the lack of any changes in efficacy in the present experiments. Narrowing the temporal range of stimuli available for attentional selection might be expected to result in a lower probability of an efficacious selection. However, selection is fast enough that even when it cannot select a long-lasting persistent representation, it still has ample time to select from the incoming stream of stimulus information. That is, the latency of selection never approaches the end of the stream. The longest latencies are found in the 18-streams condition of Experiment 2. In that condition, the shortest possible time between the target's onset and the end of the stream was 1120 ms (14 serial positions), and the mean latency was 90 ms. Evidently participants are able to select an item from the stream well before the stream ends, and thus their efficacy does not suffer even when persistence of item information is reduced. This may also explain why we were more likely to observe skew in the efficacious distribution with many streams relative to few. If representations decay faster when there are many streams, this will only affect the efficacious distribution's left tail. Attention can still select items presented after the cue, leading to skew in the distribution. However, as we have noted, the evidence for stimulus interference may reflect crowding, so this is a tentative conclusion.

4.8.2 Relation to published theories of temporal selection

No published theory of attentional selection appears capable of explaining the results of the present experiments. Most theories contain no process for attentional selection capable of resulting in a response of anything other than the target item. Those that do have such a facility include no role for endogenous attention, which is unfortunate as our results suggest it reduces latency.

As discussed, our results argue against Goodbourn and Holcombe's (2015) theory of separate processes for buffering and attentional selection from an RSVP stream. The efficacy of our participants is steady, regardless of the presence of efficacious pre-cue reports, suggesting that a single process is responsible for the pre- and post-cue efficacious selections.

Most theories of selection from RSVP streams were designed primarily to explain the attentional blink. These models do not consider errors in the timing of selection and are instead focussed on modelling the ability to report a second target soon after a first. When deployment of attention is successful, these models always predict a report of the target item (Olivers & Meeter, 2008; Wyble, Bowman, & Nieuwenstein, 2009). As such, there is no consideration of changes in the temporal properties of selection, despite the fact that the attentional blink not only affects the efficacy of selection, but also its latency (Chun, 1997; Goodbourn et al., 2016; Popple & Levi, 2007; Vul, Nieuwenstein, & Kanwisher, 2008). Here, we show that even when there is one target and thus no attentional blink, the endogenous allocation of attention to streams prior to the onset of a target stimulus can affect processing latency. The presence of efficacious pre-cue reports when there are few streams and the change in latency as the number of possible cue positions increases cannot be reconciled with these models, which predict no errors in temporal selection.

One manner in which these attentional blink models could explain this chapter's results is if attention was deployed not only in response to the cue, but in accordance with the temporal properties of the stream that a participant has learned. In this case, a pre-target item could be selected if attention was deployed in anticipation of the cue. Attention can be deployed according to temporal expectations about a stimulus (i.e. Nobre, Correa & Coull, 2007). However, participants would have to guess the location of the cue on each trial if they were deploying attention in this manner. This would reduce efficacy as the number of streams increased, as participants would be more likely to select the wrong stream. Because we don't observe this effect, we do not think temporal expectations can explain our results.

Reeves and Sperling (1986), Shih and Sperling (2002), and Grossberg and Stone (1986) modeled the data from RSVP tasks in which participants attempted to report several stimuli in succession. In these models, if a stimulus is not selected, sensory information about it is extinguished by the presentation of a subsequent stimulus at the same location. Our results demonstrate that this assumption is wrong. Stimulus representations are not terminated by the presentation of a subsequent RSVP item. They persist such that they are sometimes selected and reported even when the cue occurs after the reported stimulus.

Reeves and Sperling (1986) note that modelling stimulus representations so that they overlap with the presentation of the next item did not improve model fits. Their task involved monitoring one RSVP stream for a target and then shifting attention to another stream in order to report the letters appearing there. This presumably involves an endogenous shift of attention, whereas the present experiments used a peripheral cue which most likely induced an exogenous shift of attention. Endogenous attentional orienting is, on average, slower than exogenous orienting (Cheal & Lyon, 1991), so it may be that the speed of attentional selection was not fast enough to yield evidence of overlapping stimulus representations in their task. It remains to be seen whether Reeves and Sperling's (1986) model fits would be improved by assuming overlapping representations for a task with an exogenous cue.

Other theories of RSVP selection do provide for attentional selection of buffered items, but do not include a role for endogenous attention. Botella et al. (2001) tasked participants with selecting target words from RSVP streams. They argued that pre-target items may be selected if attention fails to focus and participants make a "sophisticated guess" about which of the features in a buffer was closest in time to the cue. They conceived of the buffer as comprising unbound feature representations activated by the items in the stream. In their theory, like ours, the time taken to process the cue is correlated with selection latency. But their model includes no endogenous mechanism that affects the cue processing time and does not assume that whole items can be reported from the buffer – whereas in our experiments this component has been shown to affect selection latency. Similarly, Vul & Rich (2010) propose a buffer of activated stimulus representations, as reviewed in the Introduction, but do not discuss a role for the allocation of attention prior to the cue.

4.8.3 Limitations

Our binomial test does not directly compare conditions, but we make an assumption that these conditions differ from each other. The binomial test allows us to infer that certain SPEs are more frequent than expected from identification failures. The test, unfortunately, does not compare across conditions, so it is possible that the different conditions do not differ in their proportion of buffered results (Gelman & Stern, 2006).

Appendix

To investigate how confidence ratings changed as a function of SPE and condition, we fit five generalised linear mixed effects models, each of which predicted the odds of making an ‘unsure’ response. The predictors varied between models. These were: the absolute value of the SPE on each trial (|SPE|), the number of streams, and an interaction term. We fit five models using lme4 (Bates et al., 2015). The models all contained random intercepts by participant. The models were:

1) Intercept only
2) |SPE|
3) Number of Streams
4) Number of Streams & |SPE|
5) Number of Streams, |SPE| and an interaction between these two variables

We interpreted the model with the lowest Bayesian Information Criterion (BIC) as the best fitting model. BIC values are presented in Table A1. The best fitting model contained |SPE| and the number of streams as predictors. Fixed effects for the model are presented in table A2. Both variables had an effect on the odds of responding “unsure”. As |SPE| increased, the odds of responding unsure increase by 0.25 (B = 0.23, z = 8.008, *p* =1.17 x 10^{-15}). Indicating that participants had some insight into whether or not their report was the target item. The odds of responding “unsure” decreased in the eight-streams condition by .36 (B = -0.45, z = 4.57, *p* = 4.8 x 10^{-6}). Indicating that participants in this condition where, in general, more sure of their answers.

Table A1

Bayesian Information Criterion (BIC) values for the generalised mixed-effects models. The best fitting model is highlighted

Model	BIC
Intercept-only	2717.36
|SPE|	2629.42
Number of Streams	2707.84

|SPE| + Number of Streams	**2616.39**
|SPE| + Number of Streams + Interaction	2624.20

Table A2

Fixed Effects Estimates from the best fitting model

	Estimate	Standard Error	z	p
Intercept	1.58856	0.44099	3.602	0.000315
Number of Streams	-0.44869	0.09811	-4.573	4.80e-06
|SPE|	0.23420	0.02925	8.008	1.17e-15

Chapter 5: Crowding and Eccentricity

Visual scenes may contain many objects. This can lead to clutter, which makes it harder to detect stimuli (Rosenholtz, Li, & Nakano, 2007) and can lead to “crowding” - interference between objects that are close together. Crowding is impaired accuracy when attempting to identify a target stimulus in the presence of flanking stimuli, despite the fact that the target stimulus is larger than the acuity threshold for its position in the visual field (Bouma, 1970). Crowding is apparent, on average, when stimuli have a centre-to-centre spacing that is less than approximately half the target’s eccentricity (Bouma, 1970). That is, the size of the crowding regions scales with eccentricity. Average identification performance improves when the target-flanker spacing is increased beyond this distance - the target is no longer crowded.

Crowding’s eccentricity scaling property has been referred to as a law (Pelli & Tillman, 2008). However, the critical spacing can be less than half the eccentricity if the target and flankers differ in contrast polarity, shape or colour (Chakravarthi & Cavanagh, 2007; Kennedy & Whitaker, 2010; Kooi, Toet, Tripathy, & Levi, 1994). Crowding has been the focus of substantial empirical and theoretical effort (For reviews, see Pelli & Tillman, 2008; Whitney & Levi, 2011). The dominant models of crowding posit that features within the crowded area are pooled, and that these pooling regions increase with eccentricity (Agaoglu & Chung, 2016; Freeman & Simoncelli, 2011; Harrison & Bex, 2015; Rosenholtz et al., 2019).

Crowding has typically been investigated in terms of spatial or featural properties, such as the effects of target-flanker similarity discussed above, with only a few investigations of its temporal properties. In their review of the crowding literature, Whitney and Levi (2011) propose empirical phenomena thought to be diagnostic of crowding, as do Pelli, Palomares and Majaj (2004). Almost all of these diagnostic criteria refer to either features of the stimuli and percept or their spatial arrangement. One exception is their suggestion that crowding's temporal properties may distinguish it from the effects of masking displays that are spatially similar to crowding displays, such as object substitution or metacontrast masking (Enns & Di Lollo, 1997).

The few existing investigations of crowding's temporal properties use different methodologies and find somewhat inconsistent results. Ng and Westheimer (2002) asked their participants to identify the location of the gap in a Landolt C flanked by four lines, each one flanking a potential position of the gap. They varied the SOA between the C and the flankers and found that the threshold gap size was highest when the flankers were presented 50 - 100 ms after the C. However, this temporal relationship is more diagnostic of object-substitution masking, as was their stimulus display. Ng and Westheimer (2002) found almost no crowding when their flanking stimuli preceded the target stimulus, a typical pattern in masking which uses flankers that do not overlap with the target (Enns & Di Lollo, 1997). When crowding is created using traditional letter triplet stimuli, flankers that appear before the target can cause crowding. This was demonstrated by Huckauf and Heller (2004), who varied the time between a target letter and two flankers and found that identification accuracy decreased as SOA decreased, even when the flankers appeared before the target.

5.1 Temporal Crowding

Some investigations of crowding have manipulated the distance between stimuli in time and named the resulting reduction in identification accuracy as temporal crowding. However, it is not clear whether this is a phenomenon different from masking. Bonneh, Sagi and Polat (2007) investigated crowding in amblyopes and controls with normal or corrected-to-normal vision. Their temporal crowding paradigm consisted of size threshold measurements for the identification of a digit presented in a single RSVP stream at fixation where the target digit was smaller than the distractor digits. They found that the size threshold for detecting the target increased with the presentation rate (2.5 Hz or 5 Hz) for strabismic amblyopes more than controls, although they do not report statistical tests of whether the threshold increases were significantly greater than zero. They named this phenomenon “temporal crowding”, and found that the extent of the threshold increase correlated with the threshold increase for spatial crowding, as measured with tumbling Es and Gabor alignment displays. Yeshurun, Rashal and Tkacz-Domb (2015) also investigated temporal crowding by varying the presentation rate of a short, three-item RSVP stream, and found that accuracy for identifying the orientation of a target letter decreased with the time between stimuli.

An issue with the notion of “temporal” crowding is that the authors do not attempt to distinguish it from masking. Masking is a reduction in identification accuracy when a briefly presented target stimulus is followed or preceded by another stimulus at the same location (Enns & Di Lollo, 2000; Kahneman, 1968). The extent of the reduction in accuracy decreases as the lag between the mask and target increases (Spencer & Shuntich, 1970). When the rate of presentation is changed in an RSVP task, as in Bonneh et al. (2007) and Yeshurun et al. (2015), the lag between successive stimuli is increased. This results in a reduction in accuracy as expected if RSVP stimuli are masking the target.

In the present experiment, we use a spatial manipulation of crowding - the size of the cue - rather than varying the presentation rate in order to produce what Bonneh et al. (2007) called “temporal crowding”. Most investigations of crowding have caused crowding by manipulating the spacing between stimuli and, as discussed above, attempts to determine the criteria that are diagnostic of crowding have relied almost entirely on spatial, rather than temporal, phenomena. It is not clear to what extent the decrease in accuracy with faster rates is a similar phenomenon to spatial crowding, rather than due to backward masking or briefer presentation of stimuli. Bonneh et al (2007) attempt to answer this with correlations between measures of temporal and spatial crowding, but this does not rule out that their temporal crowding results are due to masking.

5.2 Eccentricity and temporal processing

In addition to investigating crowding, we test for effects of eccentricity on efficacy, latency and precision. The way in which form vision changes with eccentricity is well known (see Strasburger et al., 2011 for a review) and is thought to reflect structural changes in the retina and visual cortex. Photoreceptors become sparser with eccentricity, ganglion cell receptive fields become larger, and there is less cortical area dedicated to a given area on the retina (Curcio et al., 1990; Daniel & Whitteridge, 1961; Westheimer, 2004). Typically, performance degrades as stimuli become more eccentric (but see Kehrer, 1989; Yeshurun & Carrasco, 1999). As stimuli become more eccentric, acuity decreases (Low, 1951), contrast thresholds for letter and numeral identification increase (Strasburger, Harvey, & Rentschler, 1991; Strasburger, Rentschler, & Harvey, 1994), and crowding occurs over larger regions (Bouma, 1970; Pelli & Tillman, 2008; Whitney & Levi, 2011). However, the possible relationships between temporal processing and eccentricity are less well described and the results from different tasks are inconsistent. It is well known that critical flicker fusion frequencies increase with eccentricity, such that faster flicker can be detected in peripheral vision than foveal vision (Hartmann et al., 1979; Tyler, 1987). However, the threshold for detecting a blank period between two flashes of light at the same location increases with eccentricity (Poggel & Strasburger, 2004), suggesting poorer temporal resolution in the periphery, but this may reflect the diffusion of attention over a larger area with increasing stimulus eccentricity (Poggel, Treutwein, Calmanti, & Strasburger, 2006).

Carrasco, McElree, Denisova and Giordano (2003) found evidence that visual processing became faster as stimuli became more eccentric. They investigated participants' sensitivity for identifying the orientation of a tilted Gabor patch among isoeccentric vertical distractors. To assess the time course of processing, they instructed participants to respond at different times and estimated the rate of processing by fitting exponential functions to increases in sensitivity over time. Sensitivity increased at a faster rate for stimuli presented 9º from fixation relative to 4º from fixation, and asymptotic sensitivity was higher at the more eccentric locations. Scaling the 9º stimuli in an attempt to equate the amount of cortical space associated with stimuli across eccentricities slowed processing, but did not eliminate the eccentricity differences. Participants are still faster at identifying peripheral stimuli in this paradigm when the target location is pre-cued with a transient peripheral cue (Carrasco et al., 2006).

5.3 The current experiment

The effect of eccentricity on temporal qualities of visual processing may be more complicated than eccentricity's effect on spatial vision, which typically shows a consistent degradation of spatial vision in the periphery. In some conditions, selection of more peripheral stimuli is faster even when those stimuli are cued (Carrasco et al., 2006, 2003), and temporal discrimination tasks provide conflicting evidence about temporal resolution in the periphery (Hartmann et al., 1979; Poggel & Strasburger, 2004). RSVP provides a way to test the speed of selection. Given the much-studied relationship between eccentricity and spatial vision, further investigations of the relationship between eccentricity and the temporal aspects of vision are needed. Indeed, the wider receptive fields, decreased photoreceptor density and decreased acuity associated with eccentricity predict that evidence accumulation should slow in the periphery, because eccentric visual information may be less precise and take longer to facilitate a decision.

The current experiment compares temporal selection in RSVP across different eccentricities and also investigates a potential role of crowding in our results up until this point. This is the first investigation of spatial crowding’s contribution to temporal selection from RSVP streams.

This chapter investigates crowding in RSVP by varying the spacing between the cue - a white circle - and the target stream. This experiment was prompted by the realisation that the spacings between cues and targets used in the previous experiments in this thesis were smaller than half the eccentricity of the target and thus should have caused crowding. For instance,the spacing between the centre of the cue’s line and the centre of the target letter in Chapter 4 were 0.08º, 0.15º and 0.22º at eccentricities of 3º, 7.5º and 11º, respectively. Despite this, even in the conditions where latency was most delayed, participants reported a letter on average 100 ms after the cue with an average efficacy of around 75%, even though crowding should occur up to 150 ms before or after the onset of a crowding stimulus (Huckauf & Heller, 2004). These efficacy and the latency estimates suggest that participants can report the target letter despite a centre-to-centre distance between the cue and the target letter of less than half the eccentricity. This suggests that crowding was not as strong as we might expect with letter flankers, however it may have still contributed to our results and other studies in which the stream was cued with a ring (Goodbourn & Holcombe, 2015; Holcombe et al., 2017; Vul & Rich, 2010). To test this, we increase the spacing to greater than the crowding range in this experiment in order to test for temporal changes associated with crowded and uncrowded targets.

Because it is not clear whether temporal crowding is masking, rather than crowding, the most relevant investigation of crowding’s temporal qualities to our experiment is Huckauf and Heller (2004). This study used a spatial manipulation of crowding and found impaired identification of a target letter when flanking letters were presented up to 150 ms before or after the target.

When the target stimulus is crowded in an RSVP task, identification of the target will be impaired. This might cause participants to select a letter after the target, rather than decreasing efficacy. Crowding impairs identification much more than it impairs detection (Levi, Hariharan, & Klein, 2002b, 2002a; Pelli et al., 2004) and a flanking stimulus is less crowded than the flanked target (Agaoglu & Chung, 2016) so we do not expect crowding to interfere with the detection of the cue. Instead, if the cue crowds the target, identification of the target should be impaired and participants may select a subsequent letter. This predicts a delay in selection, as identification of the crowded target is unlikely and participants select the next stimulus instead.

As for eccentricity, the predictions from theory are less clear. Increasing eccentricity may result in faster processing as predicted by Carrasco et al. (2003; 2006). However, we expect that the eccentricity benefit will result in faster accumulation of letter identity, but detection of the cue will show little benefit because stimulus detection is simpler than letter identification (Pelli, Burns, Farell, & Moore-Page, 2006). This is consistent with cue-target lag studies with static displays, which find no effect of eccentricity on the time course of attentional selection (Cheal & Lyon, 1989; Hamilton, Stark, & Coslett, 2010). Faster processing of letters, counterintuitively, results in a delay in latency. If evidence accumulation is delayed for stimuli closer to the fovea than more peripheral stimuli, then - assuming letters are processed serially - evidence accumulation about earlier letters will still be taking place when the cue is detected for less eccentric stimuli.

5.4 Method

Participants. Sixteen participants from the University of Sydney participated in this experiment. Five of these participants were graduate students, including the author of this thesis. The remaining 10 participants were psychology undergraduates participating in the experiment for course credit. The sample size was determined by the availability of participants. This experiment was not pre-registered.

Apparatus. Stimuli were presented on a Mitsubishi Diamond Pro 2070SB CRT monitor, which was 40.5 cm wide with a resolution of 1024 x 768 pixels and a refresh rate of 100 Hz. Participants viewed the monitor in a headrest from a distance of approximately 36.5 cm. The experiment was conducted in a darkened room. Fixation was not monitored using an eyetracker.

Stimuli. Participants viewed two simultaneous RSVP streams on each trial. Each stream was made up of the letters of the alphabet, with the exception of C and W, presented in a random order. The letters were white Sloan font and the cue was a white circle with a line width of 0.09°. Stimuli were presented so that their centres were either 3, 7 or 11.5° from a central fixation point, which was a white circle with a radius of 0.18°. On each trial, the two streams were presented at the same eccentricity. The streams were always on opposite sides of the fixation point such that a line drawn from the centre of one stream to the centre of the other would pass through fixation. The positions of the streams were varied between trials in order to occupy each of the six equally spaced locations around fixation equally often.

The letters were presented at a rate of 12.5 Hz. Each letter was presented for 6 monitor frames (60 ms) with a blank period of 2 monitor frames (20 ms). On a particular trial, the cue appeared in a random stream with the same duration and onset time as a random letter between the 7th and 11th items, inclusive.

The size of the letters were M scaled with the equation and values used by Strasburger (2005) to scale numerals with eccentricity:

$$S = (1 + \frac{E}{E_2})\, S_0$$

Where S is the stimulus size, E is the stimulus eccentricity, S_0 is the foveal size and E_2 is the size at which the scaled size is twice the foveal value. Our letter stimuli were never presented foveally. Previous experiments in this thesis presented 0.9° high letters at 3° eccentricity, so if we had presented letters at the fovea they would have been 0.36° high according to this function and this is the value we used for S_0. Scaling with this function then produces letter heights of 0.9°, 1.62° and 2.43° for stimuli at 3°, 7°, 11.5° of eccentricity.

There were two conditions for the crowding manipulation of our experiment. In the crowded condition - which refers to the cue-target spacings used in all previous experiments - the radius of the cue was 0.98°, 1.77° and 2.65° at eccentricities of 3°, 7° and 11.5°, respectively. The letter and crowded cue radii at 3° are the same as those used in Experiment 1 of Chapter 2 and the inner ring of Experiment 2 of Chapter 2. The sizes at 7° and 11.5° are the same as the middle and outer rings of Experiment 2 of Chapter 2.

For the uncrowded condition, we increased the radii of the cues so that the spacing between the cue's line and the centre of the target was more than half the target's eccentricity. To do this, we scaled the radius of the cue with the equation

$$Radius = 0.5 \times Eccentricity \ + 0.5 \ \times Letter\ Height \ + 1$$

The distance between a cue in the uncrowded condition and the centre of its corresponding target letter was thus one degree greater than half the target's eccentricity, and thus outside the range in which crowding is typically observed. This leads to cue radii of 2.95°, 5.31° and 7.97° at 3, 7, and 11.5 degrees of eccentricity, respectively.

Procedure. Participants were instructed to maintain fixation and report the letter that appeared within the cue on each trial. Trials of different eccentricities and cue radii were randomly intermixed. Participants performed 360 trials, which corresponds to 60 trials at each pairing of eccentricity and cue radius.

Analysis. We fit the Gaussian and gamma mixture models to each participant's data and computed model comparisons in the manner described in Chapter 2. Parameter estimates were compared using Bayesian linear models implemented in the BayesFactor Package (Morey & Rouder, 2018) in R (R Core Team, 2019). Specifically, for each kind of parameter estimate, we computed linear models in which one parameter (efficacy, latency, or precision) was the dependent variable. We computed five models for each parameter, each predicting the parameter with different dependent variables. These models were:

1. Intercept only
2. Eccentricity
3. Ring size
4. Eccentricity and ring size
5. Eccentricity, ring size and their interaction

These models were produced with the generalTestBF function from the BayesFactor package. We then computed the Bayes factor for each independent variable by calculating the inclusion Bayes factor, the ratio of the posterior model probabilities for all models containing a particular variable relative to all models without that variable (Hinne et al., 2019). Model coefficients are sampled from the posterior for the model with all predictors, although we only interpret coefficients for which there is evidence according to the inclusion Bayes factor.

5.5 Results

Excluded participants. One participant produced responses with very low efficacy (0.2 - 0.4) on all conditions. Collecting more data from this participant did not improve efficacy above 0.5 on all but one condition, so they were excluded from the analysis. A further 5 participants in the 11° eccentricity condition had precision estimates that consistently went to the upper bound of the range of values in which the mixture model fitting procedure searched for parameter estimates, even when those bounds were increased to reflect potentially wider precisions for those participants. These data were excluded from the parameter analyses. It is not clear to us why these participants data could not be fit by the model.

Model Comparisons. In all conditions, the majority of participants' SPE data were more likely under the mixture model with a Gaussian efficacious distribution than the mixture model with a gamma distribution. The results of the model comparison procedure is presented in Table 1. Because the data were most likely under the Gaussian model in the majority of cases, we used the parameter values from this model for further analysis.

Table 1.

The number of participants whose data were most likely under the different models

Cue Radius	Eccentricity of stimulus in degrees	Normal	Gamma	Neither
Uncrowded	3	14	0	1
Uncrowded	7	15	0	0
Uncrowded	11.5	14	0	1
Crowded	3	11	0	4
Crowded	7	13	0	2
Crowded	11.5	13	1	1

Efficacy

Efficacy - the proportion of responses that were not identification failures - decreased with eccentricity ($BF_{eccentricity}$ = 1600). The magnitude of this decrease was about .01 per degree of eccentricity. The crowding manipulation yielded evidence that was nearly equally well explained under the null of no effect as it was under the alternative ($BF_{crowding}$ = .64).

There was evidence against an interaction between crowding and eccentricity ($BF_{interaction}$ = .25).

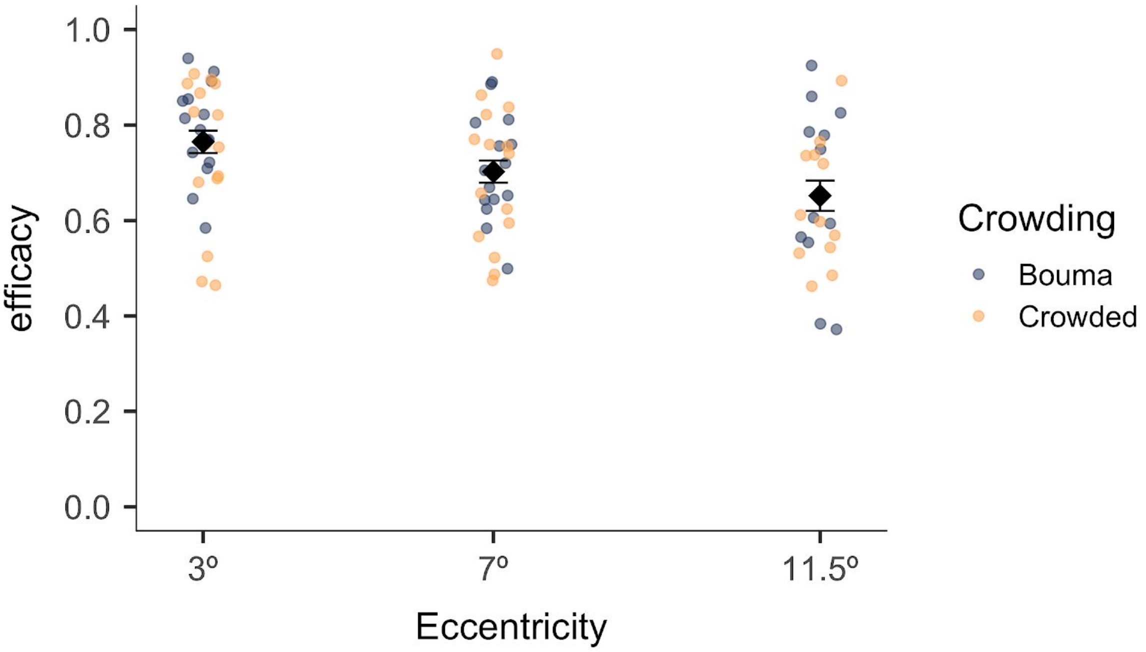

Figure 1. *Efficacy estimates for each eccentricity and crowding condition. The coloured points are the estimates for each participant. The black points and error bars are the means for each eccentricity ±SE.*

Latency. Latency was affected by crowding. When the cue was within the crowding region, latency was on average 13.8 ms later than when the cue had a radius beyond the crowding region ($BF_{crowding}$ = 5.22). The results provided little evidence of whether eccentricity ($BF_{eccentricity}$ = 0.75) or its interaction with crowding had an effect of eccentricity ($BF_{interaction}$ = 0.62).

Latencies were very low in this experiment. Mean latency was 12.6 ms (SD = 48 ms) in with the larger cue, and 30.2 ms (SD = 41.2 ms) with the crowding cue. The mean latency for the crowding cue is consistent with latencies in similar previous experiments. For instance, in Goodbourn and Holcombe's (2015) first experiment, presenting two RSVP streams with a single cue yielded a mean latency of around 25 - 30 ms, based on visual inspection of their figure. The cue radius used in that experiment (5° at 6° of eccentricity) was within the crowding region.

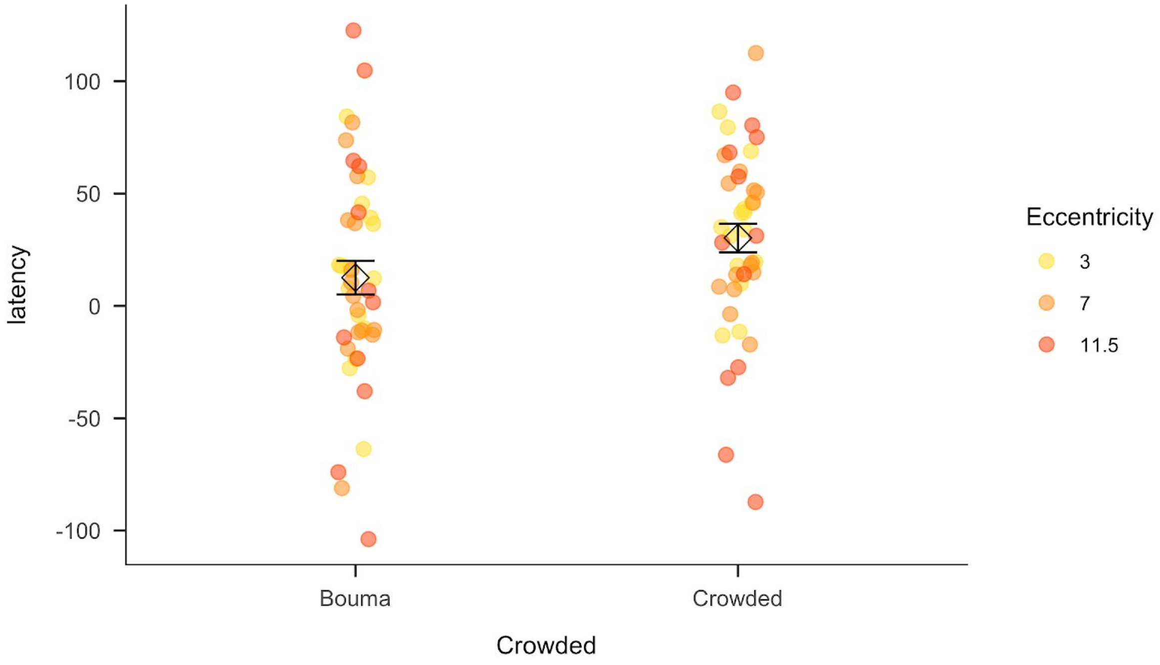

Figure 2. *Latency estimates for the crowding and eccentricity conditions. The small points are individual participants' estimates. The large points and error bars are the means ±SE.*

In Carrasco et al. (2003) and Carrasco et al. (2006), processing was faster at 9° than 4°. In an unplanned, exploratory analysis, we compared the effect of eccentricity on latency in the present study with with the rate of change observed by Carrasco et al. (2006). We compared the relationship between latency and eccentricity observed in our experiment to the change observed in Carrasco et. al.'s (2006) conjunction search condition, in which participants searched for a Gabor patch with a particular orientation and spatial frequency. We chose the conjunction search condition because our task requires detection and localisation of multiple features (the ring and the letter), more similar to the processing necessary for accurate performance in a conjunction search task than a pop-out oddball task. Selecting a cued RSVP stimulus also requires participants to orient attention to a particular location.

Conjunction search is thought to require more focused spatial attention than feature search (Carrasco & Yeshurun, 1998; McElree & Carrasco, 1999; Treisman & Gelade, 1980; Wolfe & Gray, 2007).

The speed difference between the conditions we have been discussing – 4º and 9º in Carrasco et al.'s (2006) conjunction search task – was 106 ms , which corresponds to a 21.6 ms decrease with each degree of eccentricity. This was the greatest speed difference of any condition in either Carrasco et al (2003) or Carrasco et al (2006) by 19 ms, which may bias our results against the null. But when the following analyses are conducted using the smallest speed difference, which was 17.4 ms/degree in Carrasco et al.'s (2003) feature search condition, the conclusions are the same.

The speed difference observed by Carrasco et al. (2006) is approximately 21.2 ms/degree, which corresponds to a 84.8 ms decrease in latency between our 3º and 7º conditions, or a 95.4 ms decrease between 7º and 11.5º. Using these values as the nulls for paired Bayesian t-tests demonstrated that our latency changes differed from this. In our crowded condition, the difference between 3º and 7º (M = 2.35 ms, 95% 95% confidence interval [-13.53, 18.23]) was less than the 84.8 ms (BF_{10} = 5.1 x 10^4) of the corresponding Carrasco conditions. For the participants with data in both the 7º and 11.5º conditions, the difference between the conditions (M = 3.93 ms, 95% CI [-17.40, 25.28]) was less than the 95.4 ms (paired t-test BF_{10} = 12949) of the corresponding Carrasco et al. (2006) conditions.

The pattern of results was the same in the crowded condition. The difference between 3º and 7º (M = 1.77, 95% CI [-17.00, 20.55]) was less than 84.8 ms (BF_{10} = 83028). For the participants with data at both 7º and 11.5º, the difference between the conditions (M = 18.6 ms, 95% CI [-6.04, 43.26]) was less than 95.4 ms (paired t-test BF_{10} = 890).

Thus, our analysis of the latency estimates in this experiment yielded ambiguous evidence regarding a change in latency with eccentricity. However, the results were inconsistent with the sort of latency decrease expected according to the observations of Carrasco et al. (Carrasco et al., 2006, 2003)

Precision. Precision increased with eccentricity ($BF_{eccentricity}$ = 4530), meaning that the distributions at the larger eccentricities were wider. The increase in precision was about 4.0 ms with every degree increase in eccentricity according to the linear model fit. Crowding yielded data that appeared almost as likely under the null hypothesis as under the alternative hypothesis ($BF_{crowding}$ = 0.60) and for the interaction between crowding and eccentricity, the evidence favors the null hypothesis ($BF_{interaction}$ = 0.28)

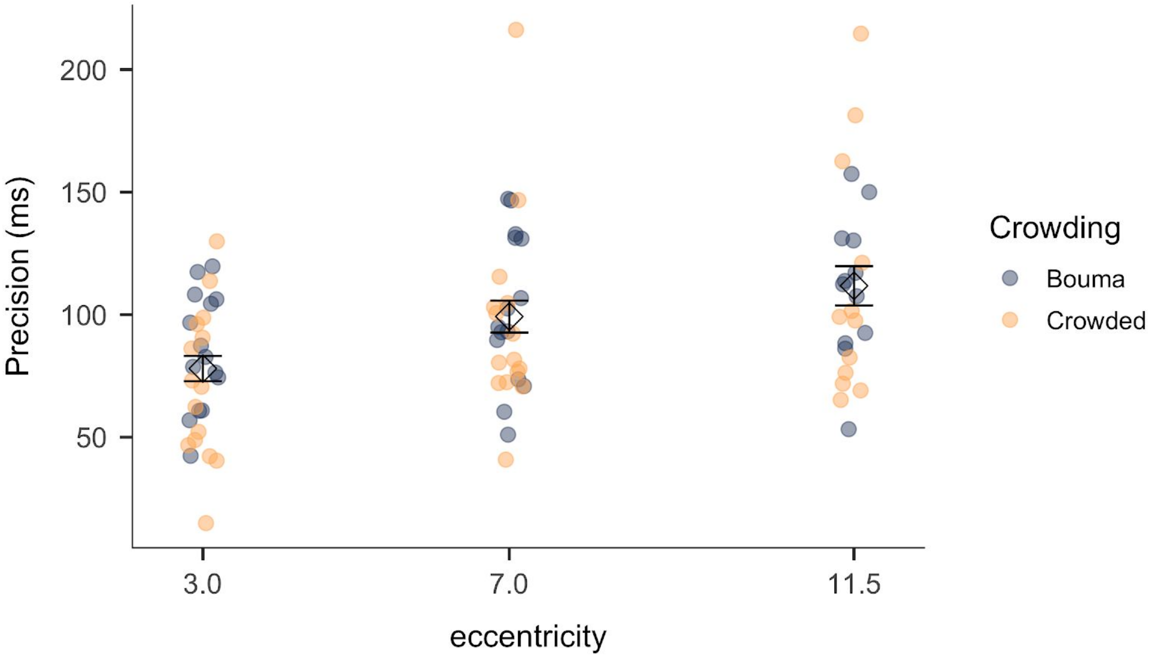

Figure 3. *Precision estimates for the eccentricity and crowding conditions.*

The Binomial Test

All participants showed efficacious reports at an SPE of -1 in the Bouma-scaled condition (*p*s ranged from 1.0×10^{-71} to 0.006). In the crowded condition, 14 of the 15 participants showed efficacious pre-cue reports at an SPE of -1 (*p*s ranged from 2.6 x 10^{-28} to 0.03); the p-value of the remaining participant and was 0.6.

5.6 Discussion

In this experiment, we found that increases in eccentricity resulted in decreases in efficacy and increases in precision. We could not make a decision about the presence or absence of a latency difference over eccentricity. However, a post-hoc analysis revealed the small or nil effect of eccentricity on latency here was smaller than the speed difference observed by Carrasco et al (2006). The only detectable effect of crowding was on latency, in which a smaller, crowding ring yielded a slight (18 ms) delay in latency.

5.6.1 Crowding

The effect of crowding was small, in the order of 18 ms for the crowding cue relative to the larger, non-crowding cue. The small effect of cue size observed here suggests that the cue sizes used in the other experiments in this thesis were not causing crowding to the extent observed by Huckauf and Heller (2004) with letter flankers. Huckauf and Heller (2004) observed impaired identification performance for crowded letters in triplets when the flanking letters were presented 150 ms before to 150 ms after the target letter, relative to identification performance for an isolated letter. When the flankers were presented with the target, identification accuracy was around 20% to 40%. This suggests that our cue does not crowd the letter at all, despite falling within the crowding region. If crowding were occurring, we would expect a much larger delay in selection because participants would report the first identifiable stimulus after the onset of the cue..

In our task, the complexity of the cue, a circle, is likely to be quite small relative to the complexity of our Sloan letter stimuli. The target stimulus is flanked on each side by the curves of the circle, rather than complex letters as in Huckauf and Heller (2004). When flankers and a target are dissimilar in shape, colour or contrast polarity, the critical spacing for identification performance is smaller than when they are similar (Chakravarthi & Cavanagh, 2007; Kennedy & Whitaker, 2010; Kooi et al., 1994). It is also the case that targets are more crowded by flanking stimuli that are more complex, relative to those that are less complex (Bernard & Chung, 2011). Because of this, the extent to which the cue crowds the target letter may be small. This means that, provided attention arrives at the cued location before the target has been replaced, participants need not select the subsequent RSVP stimulus as frequently as would be expected if the flankers were letters and crowding was strong. Attention but can sometimes select the target for identification, leading to a mean latency that is shorter than the presentation of the target stimulus.

5.6.2 Eccentricity.

As stimuli became more eccentric, the efficacy of selection decreased and precision became wider, indicating more temporal variability in selection. We were unable to make a decision about the relationship between eccentricity and the latency of selection based on our data. However, a post-hoc analysis revealed that the extent to which latency changed over eccentricity in our task was smaller than that observed by Carrasco et al (2003; 2006).

Despite M-scaling the size of our stimuli with eccentricity in an attempt to equate the cortical area dedicated to stimulus processing across eccentricities, we observed lower efficacy for more eccentric stimuli. That is, the probability of misidentifying the cue or a selected letter decreased with increasing eccentricity. The observation that M-scaling stimuli in order to account for cortical magnification fails to equate identification performance has been made before. Strasburger et al. (1994) M-scaled numerals in an identification task and observed that this was not enough to equate contrast thresholds for identification across eccentricities. It is possible that if a steeper M-scaling function were used, efficacy would be equated across eccentricities. However, the theory that cortical magnification can explain performance across eccentricities is beset by failures to scale and contradictory evidence, despite its popularity (Strasburger et al., 2011).

The variability in the timing of selection of an RSVP item when the cue is detected and an item is identified- an efficacious response - was greater for more eccentric stimuli in our experiment. This widening of precision with increasing eccentricity may reflect the increase in processing noise with eccentricity. Hess, Baker, May, & Wang (2008), by measuring contrast thresholds for letter identification in the presence of luminance noise, found evidence that more eccentric stimuli are subject to more internal noise, although their stimuli were not scaled with eccentricity. Internal noise is variation that differs with each presentation of a stimulus, which for these experiments could manifest as trial-to-trial variation in processing of the cue, resulting in variation of its detection time.

We failed to replicate Carrasco et al.'s (2003; 2006) findings of faster letter processing with increased eccentricity. This result must be taken with a grain of salt, because our precision was not great enough to lead to a decision in favour of either the null or the alternative hypothesis and the comparison with Carrasco et al.'s (2006) results was post-hoc. We did not run any unreported tests before deciding on the comparison with Carrasco et al.'s (2006) results, but it is possible that observing the data led us - unconsciously - to a test that we knew would detect a difference (Gelman & Loken, 2013).

Chapter 6: Discussion

Attention, the ability to selectively process visual information based on location or feature, is an important determinant of human visual performance (Carrasco, 2011). This thesis was an investigation of the timing properties of attention as measured by RSVP - in which participants must report one stimulus from a sequence. This paradigm was designed to mimic the stream of information presented to the visual system in natural viewing. Here we showed that attention has access to sensory memory when stimuli are presented in rapid succession, despite the masking inherent in such displays. This is inconsistent with certain published theories of RSVP. For example, some models assume that stimulus representations are terminated when a new stimulus is presented at the same location (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002). Others assume that some information persists, but that attention either does not sample it in a way that would result in temporal errors or that the persistent information consists of unbound features (Botella, Barriopedro, & Suero, 2001; Olivers & Meeter, 2008; Vul & Rich, 2010; Wyble, Bowman, & Nieuwenstein, 2009). Our results show that such assumptions may be wrong. Attention can select persistent information about stimuli that are masked and no longer presented, suggesting that visual resources can be flexibly applied to sensory memory as well as the incoming stream of visual information.

Typically, attention is studied by varying the lag between an attentional cue and a target stimulus, which has provided reliable information about the relative speed of different kinds of attentional orienting (Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989). However, such studies reflect a cumulative distribution of attentional arrival times and do not provide trial-level information about the timing of attention in response to a stimulus. Using RSVP, rather than cue-target lag, allows researchers to estimate the timing of attention in a manner that mimics the dynamism of visual scenes

(Potter & Levy, 1969; Reeves & Sperling, 1986; Weichselgartner & Sperling, 1987). However, inferences about attention's timing from RSVP data are qualified by the possibility of *identification failures*, responses that are misidentifications of a selected letter or guesses that occur when participants miss the cue. Identification failures can result in responses from any time within a particular trial, rather than being informed by the occurrence of the cue, and thus they impair our ability to draw inferences about attentional selection based on the temporal distribution of SPEs.

Mixture modelling the distribution of SPEs (Goodbourn & Holcombe, 2015) allows us to account for identification failures and draw inferences only about *efficacious* reports, those responses that are informed by the timing of the cue. This leads to a surprising finding under some conditions - the distribution of efficacious reports seems to include items from before the time of the cue (Goodbourn & Holcombe, 2015; Holcombe, Nguyen, & Goodbourn, 2017). This is surprising because many theories of attentional selection, even those that attempt to explain RSVP, assume that attention is triggered by the cue and that it cannot select information about stimuli that are no longer presented (Chun & Potter, 1995; Olivers & Meeter, 2008; Weichselgartner & Sperling, 1987; Wyble et al., 2009). Goodbourn and Holcombe (2015) argued that efficacious pre-cue reports were evidence for a perceptual buffer, a process distinct from attentional selection from the incoming stream of visual information. This thesis investigated the presence of buffering, as evidenced by pre-cue reports that were not identification failures.

6.1 The mixture model measures the temporal aspects of selection

In Chapter 2, we outlined our analyses for the SPE data. Our mixture modelling technique extended Goodbourn and Holcombe's (2015) approach to analysing SPE data and investigated the presence of skew, which we predicted would be present in efficacious selections when items were selected using an attention shift, rather than the buffer. To compare the evidence for skewed, rather than symmetric (Goodbourn & Holcombe, 2015; Holcombe et al., 2017), efficacious distributions we fit two models to each participant, one with a skewed, gamma distributed efficacious distribution, and one in which the efficacious distribution was a Gaussian.

In that chapter, we showed that selection from the RSVP streams was temporal, rather than based on items, and that the seeming-symmetric efficacious distributions observed in earlier studies were not due to slow presentation rates which provide coarse information. Mixture modelling applied to the data from a two-stream RSVP task with presentation rates that ranged from six letters/second to 24 letters/second revealed that the mixture models estimates of the timing properties of efficacious selection were truly temporal and were not affected by changes in the probability of an efficacious response. As the presentation rate increased, efficacy dropped, but the precision and latency of selection were consistent with the null hypothesis of no change across rates. Increasing the rate of presentation did not produce data that were more likely to be skewed.

Botella and Eriksen (1991) observed a shift from a symmetric to a post-target SPE distribution when they increased the presentation rate of a single RSVP stream. However, in our task participants produced data consistent with a Gaussian efficacious distribution across all presentation rates. We attribute the difference between our results and Botella and Eriksen's (1991) to differences in analysis. Botella and Eriksen (1991) analysed their data in terms of item serial position. However, If selection has a fixed positive latency, increasing the presentation rate of an RSVP stream will result in selection of later items - evidence that Botella and Eriksen (1991) interpret as a change in the shape of the SPE distribution. Consistent with this claim, we observed evidence that latency, measured in milliseconds, did not change across presentation rates

Our results also indicated that the presence of SPE distributions in previous studies with a seemingly Gaussian component, rather than skew, were not due to coarse temporal information produced by presentation rates of eight to 15 letters/second, which may obscure skew (Botella & Eriksen, 1991, 1992; Gathercole & Broadbent, 1984; Goodbourn & Holcombe, 2015; Holcombe et al., 2017; Lawrence, 1971). Likewise, a reanalysis of Goodbourn and Holcombe's (2015) single-target condition indicated that 23 of the 26 participants in that experiment produced data that were most consistent with a Gaussian efficacious distribution rather than the skewed gamma distribution.

In Chapter 2 we also outlined our binomial test, which allows us to investigate whether a participant's responses at a particular SPE are more frequent than predicted under identification failures. This allows us to test for the presence of buffered information while accounting for the fact that some trials are likely to yield identification failures, which can result in reports of items from before the cue. Our test is conservative with regards to efficacy, so it is biased towards identification failures slightly. Applying this analysis to data from previously published experiments (Goodbourn & Holcombe, 2015; Vul, Nieuwenstein, & Kanwisher, 2008) yields evidence that reports of items presented one item before the cue are more frequent than can be explained by identification failures - evidence of buffered information.

6.2 Endogenous and Exogenous cueing

Once we observed evidence that our latency and precision estimates are temporal, we began to assess the time course of selection. In Chapter 2, we compared selection with exogenous and endogenous attentional cues. Comparisons of these kinds of cues are common in cue-target lag investigations of attention's time course and have played an important role in the theoretical distinction between endogenous and exogenous sources of attentional orienting (i.e. Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama &

Mackeben, 1989; Posner, 1980). However, these comparisons are rare in the RSVP literature. When endogenous and exogenous attention are compared in the RSVP literature, the data are not mixture modelled and estimates of attention's temporal properties are likely affected by the presence of identification failures (Weichselgartner & Sperling, 1987).

Comparing a central, endogenous cue with a peripheral, exogenous cue, we found that endogenous cueing resulted in lower efficacy and wider precision than exogenous cueing. We also replicated the delay in latency associated with endogenous cueing relative to exogenous cueing commonly found in cue-target lag studies (Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989). We believe that the differences in latency and precision in this study reflect the need for participants to interpret an endogenous cue and voluntarily shift attention to the cued location. This process is longer and more temporally variable than selection based on an exogenous cue, which occurs at the cued location and requires little, if any, interpretation.

6.3 Is there a buffer?

In a dynamic visual scene, representations of information at a given location in the visual field may be replaced, either due to the observer's movements or the movement of something in the scene. This replacing of information is thought to result in masking, in which accuracy for identifying a target stimulus is impaired when another stimulus is presented at the same location shortly before or after the target (Enns & Di Lollo, 2000; Kahneman, 1968). Masking limits the time frame during which stimulus information can be selected by attention. Without masking, information about stimuli may persist in sensory memory and may be available for attentional selection even when stimuli are no longer presented, but upon presentation of a mask this information is no longer available (Averbach & Coriell, 1961; Pinto, Sligte, Shapiro, & Lamme, 2013; Sligte, Scholte, & Lamme, 2008; Sperling, 1960). RSVP mimics the dynamic stream of visual information and masking present in a visual scene by presenting a sequence of stimuli quickly at one spatial location, and models of selection from RSVP assume that masking either terminates sensory memory (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002), or that persisting information cannot be accessed by attention (Olivers & Meeter, 2008; Wyble et al., 2009).

Goodbourn and Holcombe (2015) found evidence that masking may not prevent the selection of information from sensory memory in dynamic visual scenes. In a task where participants had to select cued items from two RSVP streams, Goodbourn and Holcombe (2015) observed efficacious distributions with latency and precision estimates that suggested that some efficacious selections came from before the time of the cue. The presence of these efficacious pre-cue reports indicates selection from sensory memory of RSVP letters, which Goodbourn and Holcombe (2015) referred to as a buffer. The apparent use of buffered information suggests that despite visual representations being subject to masking in naturalistic viewing, attention can access sensory information about stimuli that are no longer present. However, models of RSVP assume that such selection from sensory memory cannot occur (Chun & Potter, 1995; Grossberg & Stone, 1986; Olivers & Meeter, 2008; Reeves & Sperling, 1986; Shih & Sperling, 2002; Wyble et al., 2009).

Goodbourn and Holcombe’s (2015) theory of buffering distinguishes buffering from attentional engagement with the stream. The buffering theory assumes that representations are activated in the buffer by incoming stimuli, regardless of whether a cue has been presented. Selection from the buffer is achieved by binding the representation of the cue with the activated representation of a letter from the RSVP stream in Goodbourn and Holcombe’s (2015) theory, which can sometimes result in the efficacious report of an item from before the time of the cue. Attentional sampling from the stream, on the other hand, is thought to occur when the cue triggers attention to engage with the incoming stream of visual information. Before the onset of the cue, selection does not occur (Chun & Potter, 1995; Olivers & Meeter, 2008; Weichselgartner & Sperling, 1987; Wyble et al., 2009).

In Chapter 4 and 5, we investigated the evidence for buffering by investigating the circumstances where participants made efficacious reports of stimuli from before the cue. Unlike previous papers that had suggested that buffered representations can be selected, we directly tested for efficacious reports of pre-cue stimuli using our binomial procedure. Chapter 4's Experiments 1 and 2 and the experiments of Chapter 5 demonstrated efficacious pre-cue reports when participants must report a cued item from one of two simultaneous RSVP streams. As we increased the number of simultaneous streams in Experiments 1 and 2 of Chapter 4, efficacious pre cue reports were less evident, the latency of efficacious selections was delayed, and the precision of efficacious selections narrowed. Efficacy did not change as the number of streams increased in either experiment, which suggested to us that pre- and post-cue reports are the result of the same process, rather than separate buffering and attentional selection processes as assumed by Goodbourn and Holcombe (2015). Experiment 3 in Chapter 4 provided evidence that the delay in selection was due to participants dedicating attention to the potential locations of the cue prior to its presentation. In that experiment, we kept the number of streams constant and manipulated participants' knowledge about the location of the cue with a pre-cue. Selection was delayed when there were eight possible locations of the cue, relative to two, suggesting that the changes in the temporal dynamics observed in Experiment 1 and 2 of Chapter 4 were due to participants attending to potential cue locations.

The results of Chapter 4 changed our theory of the role of buffering in visual performance. Initially, we thought of the buffer and attentional selection as separate processes, like Goodbourn and Holcombe (2015), but it is hard to reconcile this with the lack of an efficacy change as the number of streams increased. One would reasonably expect buffering and attentional selection to have different probabilities of reporting a letter from around the time of the cue, but we observed evidence that efficac~~y~~ did not change as the number of streams increased. Instead, we now believe that efficacious pre-cue reports are present when attentional engagement is fast enough that it may select a pre-cue letter representation before that representation decays.

Under our account, participants spread attention over all the streams presented on a particular trial prior to the onset of the cue. When there are fewer streams, more attentional resources are dedicated to a given stream and this speeds selection upon presentation of the cue. Diffusing attention over a larger area is associated with decrements in performance, such as slower reaction times for detecting a target stimulus (Castiello & Umiltà, 1990, 1992) and increased thresholds for detecting a temporal gap between stimuli (Poggel, Treutwein, Calmanti, & Strasburger, 2006). Because of this, we believe that when participants view multiple simultaneous RSVP streams, they devote some proportion of attention to each one. As the number of streams increases, a smaller proportion of attention is dedicated to each stream and this delays selection based on the cue. We do not know whether this is due to a delay in cue detection, like the delay in stimulus detection in Castiello and Umiltà (1990; 1992), or a delay in the orienting of attention to the cued location.

In order for this theory to account for efficacious reports of items from before the cue, a stimulus' representation must persist beyond the onset of the following stimulus at the same location. The phenomenon of attentional selection from visual sensory memory stores is well-established in research on iconic memory and fragile memory (Averbach & Coriell, 1961; Coltheart, 1980; Pinto et al., 2013; Sligte et al., 2008; Sperling, 1960) and it has been incorporated into theories of selection in RSVP (Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002). However, the literature on sensory memory states that it is erased by post-masking when stimuli are presented at the same location in visual space. Under these theories the memory process should be eliminated in RSVP, where subsequent items mask previous items from the same stream. However, recently it has been observed that selection from sensory memory is possible even in the presence of a post mask (Smith, Mollon, Bhardwaj, & Smithson, 2011; Smithson & Mollon, 2006), and neural decoding studies indicate that information about stimulus identity in RSVP persists during the presentation of a subsequent item in the same location (King & Wyart, 2019; Marti & Dehaene, 2017). Our results add to this growing body of literature on the resilience of sensory memory to masking. We observed that the efficacy of selection was unchanged by the number of simultaneous streams, even though pre-cue selections became less frequent. This suggests that the process involved in selecting items from before the cue is the same as that involved in selecting items from after the cue. In both cases, attention selects stimulus representations, and the pre-cue reports indicate that sensory memory is the source of these selections even in the presence of masking.

Our theory thus differs from that of Goodbourn and Holcombe (2015). They assumed that representations of letters and the cue were activated in a buffer and that one representation was bound with the cue for report. This is distinct from the cue triggering sampling from the stream, which we predicted would dominate at higher numbers of streams (https://osf.io/7hkgd). Our results prompted a theory that differs from this buffering and binding process. The buffer is retained, in the form of a memory story that is resilient enough to masking to provide information for selection. However, selection occurs not through a process of binding the cue with a letter representation, but instead attention is triggered by the cue and if the attentional response is fast enough it can select information from sensory memory about a pre-cue letter.

The idea that masking terminates the representation of stimuli in dynamic displays is long standing. Sperling (1960) provided early evidence that this was the case, and 35 years later, Chun and Potter (1995, pp. 109) stated that each RSVP stimulus “eliminates the previous item from sensory storage”. The experiments reported in this thesis provided evidence that this assumption is wrong. By designing the first direct test of efficacious pre-cue reports, we demonstrated that sensory memory in dynamic displays persists despite masking and can be selected by attention, provided attention is fast enough. Current theories of RSVP cannot account for these results. As discussed in Chapter 4, several theories of selection in RSVP assumed that sensory information is inaccessible in the presence of post-masking, so these results falsify those theories by demonstrating efficacious pre-cue selections (Chun & Potter, 1995; Grossberg & Stone, 1986; Reeves & Sperling, 1986; Shih & Sperling, 2002).

While Chun and Potter (1995) stated that each RSVP stimulus masks the stimulus that came before it, one element of their theory may explain the results we observe here. In order to explain the attentional blink, Chun and Potter (1995) proposed a two-stage model in which RSVP stimuli are first processed for feature detection in a preattentive manner and - if a target stimulus is detected - are consolidated with attentive processing. Their preattentive stage contains a brief store of conceptual information - such as a stimulus’ category and semantic associations - known as conceptual short term memory (CSTM; Potter, 1993). CSTM contains conceptual information accessed even when stimuli are presented at fast (8 to 10 Hz) presentation rates, but it is quickly forgotten. CSTM can explain priming effects in the attentional blink, in which accuracy for identifying the second target in an RSVP stream is increased when that target is preceded by a semantically related distractor or target (Maki, Frigen & Paulson, 1997; Shapiro et al., 1997). These effects can be explained if we assume that semantic information about RSVP stimuli is accessed even with brief, masked presentations, and this information persists so that if a semantically-related target is presented before it decays, it benefits from this activation.

CSTM is thus a kind of buffer, in which stimulus information persists and can be used for selection. However, conceptual relationships between representations in CSTM are critical for information to persist (Potter, 1993). Without conceptual structure, information decays quickly, as indicated by observations that recall for targets in RSVP streams of words is more accurate when the stream's words form a coherent sentence relative to when they do not (Potter, Nieuwenstein and Strohminger, 2008). The letters that we use in this thesis have little conceptual structure, so under this theory they would decay quickly. If the decay rate for items with little conceptual structure is long enough that one item's activation persists beyond the next, then CSTM could explain the selection of buffered items that we observe.

Other theories allow for stimulus information that persists after the next stimulus, but none of these contain a role for the allocation of attention prior to the onset of a cue, but we demonstrate here that attending to the streams prior to the onset of the cue affects the speed with with attention engages with the cued stream (Botella et al., 2001; Olivers & Meeter, 2008; Rusconi & Huber, 2018; Vul et al., 2008; Wyble et al., 2009). These theories need to be modified in order to account for the results of this thesis.

6.3 Skew and symmetry in temporal selection

In conditions with few streams and a ring cue, we consistently observed data that were better fit by the mixture model with a symmetric efficacious distribution rather than skew. This was surprising, because we assume that selection is the result of attention, triggered by the cue. An attentional episode triggered by the cue in this manner should impart a positively-skewed efficacious component to the SPE distribution, because selection has a lower bound imposed by the presentation of the cue. This assumption was formalised in our gamma mixture model, which used a skewed efficacious distribution. The seeming absence of skew in the efficacious distributions does not seem to be due to the coarse grain of the temporal information that the RSVP paradigm provides. There were no systematic increases in the evidence for skew when the presentation rate increased in Chapter 2 - most participants in that experiment produced data that favoured the symmetric Gaussian mixture model.

Symmetric and skewed SPE distributions were previously observed by several researchers interested in how the feature and identity of letter or numeral RSVP stimuli were processed (Botella, 1992; Botella & Eriksen, 1992; Broadbent & Broadbent, 1986; Gathercole & Broadbent, 1984; McLean, Broadbent, & Broadbent, 1983). These data were not mixture modelled and inferences about symmetry were drawn by comparing the percentage of the SPE distribution before and after the target. Distributions made up predominantly of post-cue (i.e. not symmetric) stimulus reports, like those observed in Lawrence (1971), were interpreted as evidence for serial processing of the cue feature (i.e. a particular stimulus' colour) before processing stimulus identity (Broadbent, 1977). The symmetry of SPE distributions, as measured by comparing the proportion of the SPE distribution before and after the target, was interpreted as evidence of parallel processing of feature and identity information, and changes from symmetric to post-target SPE distributions were interpreted as strategic shifts between serial and parallel processing modes (Botella & Eriksen, 1992; Broadbent & Broadbent, 1986; Gathercole & Broadbent, 1984; McLean et al., 1983). However, as we demonstrate in Chapter 2, these changes are consistent with a symmetric distribution fixed positive latency, rather than a change from symmetric to skewed SPE distributions.

We mixture modelled the SPE distributions, compared the evidence for skew using a Bayesian model comparison procedure and observed consistent evidence for symmetric efficacious distributions (the Gaussian model was favoured) with few streams and a ring cue (Chapter 4 and Chapter 5). We still observed symmetry when selection was cued with a central cue (Chapter 3), a condition which we presume resulted in an attention shift from the centre of the display to the cued stream, because participants would need to attend centrally to detect the cue and identify its location relative to fixation. A peripheral cue, on the other hand, provided the location of the cued stream and participants appeared to attend to the streams prior to the cue's onset. Three of the six participants in the central cue condition produced data that were more likely under the Gaussian model. The presence of a central, rather than a peripheral, cue had the effect of delaying selection and increasing its variance, but the data for half the participants were more likely under the symmetric Gaussian model than the skewed Gamma model.

Symmetry does not appear to allow us to diagnose the presence of triggered attention shifts. We believe this is due to persistence in the representations of RSVP stimuli. In Chapter 4 and 5, we consistently observed efficacious pre-cue reports, and the results in Chapter 4 lead us to believe that attentional selection is responsible for these reports. We explain the presence of selection from before the onset of the cue with the informational persistence of stimulus representations. Similarly, if there is information about stimuli that are no longer presented available for selection, then the decay of these representations may explain symmetry in the SPE distributions by allowing the possibility of attentional selection of stimulus representations that are no longer presented, contributing to the left tail of the SPE distributions.

We only observed skew in the distributions when there were many streams, but because we did not observe a change in efficacy for these skewed distributions, we do not believe that these reflect an attention process distinct from that which produces efficacious pre-cue selections. One reason we may have observed skew in these conditions relative to those in which there were few streams is that the presence of multiple simultaneous RSVP streams lead to competition between stimulus representations. With many streams, there is increase competition between stimuli, resulting in stimulus-induced activity returning to baseline quicker than with fewer streams.

6.4 Comparison between cue-target lag and RSVP

Traditionally, the time course of attentional selection has been inferred from studies in which the lag between an attentional cue and some target stimulus is varied and participants' accuracy for responding to the stimulus is measured (Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Posner, 1980). Accuracy typically increases as lag is increased from 0 to 100 - 300 ms, depending on whether attention is cued with a central (endogenous) or peripheral (exogenous) cue. The lag at which accuracy peaks (in the case of exogenous attention) or asymptotes (in the case of endogenous attention) provides researchers with an estimate of the earliest time at which attention provides maximum benefit for target processing. However, as we have discussed in Chapter 1, this provides cumulative information about the arrival of attention. If a set of trials with a particular lag is significantly more accurate than trials with a shorter lag, then we may infer that attention arrived at or before the longer lag, but was unlikely to arrive at or before the shorter lag. This inference is complicated somewhat in studies with a peripheral cue, because long cue lags can result in accuracy that is worse than that around 100 ms, indicating that attention has disengaged from the cued location on some proportion of trials. In this case we do not know when attention arrived and when it disengaged.

We used RSVP and mixture modelling in order to produce a distribution of attentional arrival times from trials that were efficacious. RSVP allows (quantised) trial-level inferences about the timing of attention based on the stimulus reported by a participant on each trial. Identification failures - responses from trials in which the cue or the selected letter were misidentified - can result in reports of items from any point in time in a trial, which contaminate these estimates of attention's timing. Our mixture modelling processing accounts for this and allows us to analyse the temporal qualities of efficacious selections. Rather than tracing the cumulative distribution of attention across lags, we estimate the temporal distribution of attention as demonstrated by efficacious reports.

In Chapter 3, we compared the distribution of selections from one of six streams produced by central and peripheral cues. Comparing these sorts of cues in cue target-lag studies has provided important evidence for the theoretical distinction between endogenous and exogenous attention (Cheal & Lyon, 1991; Müller & Rabbitt, 1989; Posner, 1980). We replicated the observation that central cues yield slower attentional orienting than exogenous in our latency estimates (Cheal & Lyon, 1991; Müller & Rabbitt, 1989). However, our latency estimates were much shorter than those typically observed in cue-target lag studies.

Our central cue estimates had a mean of 212 ms, shorter than the ~300 ms typically observed with central cues. Differences in latency between experiments could potentially be attributed to the use of different central cues - we used a 0.07° pixel at fixation whereas Cheal and Lyon (1991) used a 0.8° arrow and Müller and Rabbit (1989) use an arrow of unknown size. However, given that our central cue was smaller than Cheal and Lyon’s cue and hard to detect (mean efficacy with this cue was .61), the idea that differences in the cue would explain our faster latency seems implausible.

Peripheral cueing led to shorter latency estimates than the 100 - 120 ms peak times with peripheral cues observed in cue target lag studies with less than 18 streams in Chapters 3, 4 and 5 (Cheal & Lyon, 1991; Folk, Remington, & Johnston, 1992; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Posner & Cohen, 1984; Remington, Johnston, & Yantis, 1992). Latency increased as we increased the number of streams in Chapter 4 up to a mean of 90.9 ms, close to the range of values from cue-target lag studies. Perhaps the difference between our latencies and the time of peak accuracy in a cue-target lag study can be attributed to the fact that such studies often use a fixed number of items. However, Nakayama and Mackeben (1989) varied the number of items in their search arrays and found no effect on the peak timing of attention, which argues against this explanation.

We believe that the shorter latency estimates can be explained by the persistence of visual representations and the allocation of attention prior to the onset of the cue. These qualify any attempt to use RSVP in order to investigate the time course of attentional selection. We observed efficacious pre-cue selections in Chapters 4 and 5, indicating that visual representations persist and can be selected by attention based on the cue. This means that the item reported on a particular trial could be an underestimate of the time taken for attention to orient in response to the cue on that trial. Attention may select an item that is no longer presented, making selection appear faster than it is. This is not only the case for efficacious pre-cue selection. Post-cue efficacious selections may result in selection of an item that is no longer presented.

Similarly, the allocation of attention to the streams prior to the onset of the cue speeds the attentional response to the cue. When there are few streams and thus few potential cue locations, participants can dedicate more attentional resources to the cue locations and this speeds their response to the cue. This issue does not occur in cue-target lag studies, where participants typically see the cue prior to the onset of any other stimuli and cannot allocate their attention to a particular location (other than the display) prior to this in a way that would speed attention.

6.5 Future Directions

We present evidence for attentional allocation prior to the onset of a cue and efficacious reports of items that were presented prior to the cue. Some models of selection from RSVP streams assume that attention is not engaged with the stream until it is triggered by the cue (Chun & Potter, 1995; Olivers & Meeter, 2008; Weichselgartner & Sperling, 1987;

Wyble et al., 2009). Many of these models are designed to explain the attentional blink (Chun & Potter, 1995; Olivers & Meeter, 2008; Wyble et al., 2009). Current models of the attentional blink assume that the blink's time course represents encoding time into working memory (Wyble et al., 2009) or the time course of inhibition in response to detecting a distractor (Olivers & Meeter, 2008). However, there is evidence that the blink is affected by the timing of attention due to its spatial diffusion. This may call the assumptions that the blink reflects the dynamics of encoding or suppression into question.

Diffusing attention over many streams may delay selection and extend the blink. There are several attentional studies that use multiple RSVP streams in which the first and second targets can appear in different streams. Several of these have investigated how selection of one target affects the ability to switch attention to another stream (Jefferies, Ghorashi, Kawahara, & Di Lollo, 2007; Kristjánsson & Nakayama, 2002; Lunau & Olivers, 2010; Shih, 2000). None test whether the number of simultaneous streams changes the blink, but two studies provide evidence that the distribution of attention prior to the cue may extend the blink. Jefferies et al. (2007) presented participants with two RSVP streams and observed that switching from one stream to the other was impaired when participants did not know where the first of two targets would appear, so that they had to attend to both streams rather than just one. Lunau and Olivers (2010) report evidence that when two successive targets appeared in one of 27 simultaneous streams, detection of the second target was impaired from 200 to approximately 800 ms after the first target. This is much longer than the 200 to 500 ms impairment typically observed in attentional blink studies with a single stream (Martens & Wyble, 2010). Shih (2000) presented their participants with two simultaneous streams and also observed an extended blink. However, Kristjánsson and Nakayama (2002) found what appeared to be a typical time course for the blink when the targets appeared in one of eight simultaneous streams, so the effect may not be as robust as the latency increases observed in Chapter 4 of this thesis.

Another potential extension of the work in this thesis is a test of Goodbourn and Holcombe's (2015) pseudoextinction effect - an advantage for making an efficacious report in the left or superior stream of two simultaneously cued RSVP streams. Goodbourn and Holcomeb (2015) hypothesised that buffered information was the source of pseudoextinction. They argued that stimulus representations persist in the buffer, and a serial tokenisation process that begins with the left or superior stimulus consolidates stimulus information into working memory. Our results suggest that the buffer may not operate at high numbers of streams, and thus pseudoextinction may not be present under similar conditions.

The evidence for the lack of a buffer with high numbers of streams comes from the skew present in the efficacious distributions in these conditions in Chapter 4. Skew is important here, rather than the increased delay we observed, because the delayed selection may still selected a persistent representation, but not one that occured before the cue. Skew, however, suggests that suggests that earlier items from the stream do not persist long enough to be selected. This moves the lower bound on selection closer to the modal time of selection, but the right tail of the distribution is preserved due to the incoming stream of RSVP stimuli. With few streams, there is symmetry in the efficacious distribution, suggesting persistent representations.

If the buffer requires persistent information, as Goodbourn and Holcombe (2015) theorise, then conditions in which information does not persist may not demonstrate a pseudoextinction effect. By increasing the number of RSVP streams and cueing two of them simultaneously, we can test this. Participants with skewed efficacious distributions will show no pseudoextinction effect if this is the case.

6.6 Summary

The visual world is rich and dynamic. However, human visual resources are limited and we must flexibly apply them to process important aspects of the stream of incoming visual information. RSVP allows us to conduct controlled laboratory investigations of the temporal dynamics of attentional selection. Theorists typically assume that masking - interference between two stimuli presented sequentially at the same spatial location - eliminates sensory information about RSVP stimuli that are no longer presented, or attenuates it such that it is not useful. Here, we demonstrate that stimulus representations in RSVP persist beyond the presentation of a subsequent stimulus at the same location and that attention may select these representations in response to a cue. This violates assumptions about masking in RSVP and provides valuable information about how attention samples the visual world. Attention has access not only to the incoming stream of visual information, but briefly persisting sensory memory as well.

References

Agaoglu, M. N., & Chung, S. T. L. (2016). Can (should) theories of crowding be unified? *Journal of Vision*, *16*(15), 10–10. https://doi.org/10.1167/16.15.10

Averbach, E., & Coriell, A. S. (1961). Short-Term Memory in Vision. *Bell System Technical Journal*, *40*(1), 309–328. https://doi.org/10.1002/j.1538-7305.1961.tb03987.x

Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual Clutter Causes High-Magnitude Errors. *PLOS Biology*, *4*(3), e56. https://doi.org/10.1371/journal.pbio.0040056

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, *55*(4), 412–428.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, *8*(8), 1110–1116. https://doi.org/10.1038/nn1501

Beck, D. M., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision*, *7*(2), 19–19. https://doi.org/10.1167/7.2.19

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*(10), 1154–1165. https://doi.org/10.1016/j.visres.2008.07.012

Bernard, J.-B., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target–flanker similarity. *Journal of Vision*, *11*(8), 1–1. https://doi.org/10.1167/11.8.1

Bialek, W. (1987). Physical Limits to Sensation and Perception. *Annual Review of Biophysics and Biophysical Chemistry*, *16*(1), 455–478. https://doi.org/10.1146/annurev.bb.16.060187.002323

Bonneh, Y. S., Sagi, D., & Polat, U. (2007). Spatial and temporal crowding in amblyopia. *Vision Research*, *47*(14), 1950–1962. https://doi.org/10.1016/j.visres.2007.02.015

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. https://doi.org/10.3758/BF03330440

Botella, J., Barriopedro, M., & 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.

Botella, J., & Eriksen, C. W. (1991). Pattern changes in rapid serial visual presentation tasks without strategic shifts. *Bulletin of the Psychonomic Society*, *29*(2), 105–108.

Botella, J., & Eriksen, C. W. (1992). Filtering versus parallel processing in RSVP tasks. *Perception & Psychophysics*, *51*(4), 334–343. https://doi.org/10.3758/BF03211627

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*(5241), 177.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review*, *114*(1), 38.

Broadbent, D. E. (1977). The hidden preattentive processes. *American Psychologist*, *32*(2), 109–118. https://doi.org/10.1037/0003-066X.32.2.109

Broadbent, D. E., & Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113.

Broadbent, D. E., & Broadbent, M. H. P. (1986). Encoding Speed of Visual Features and the Occurrence of Illusory Conjunctions. *Perception*, *15*(5), 515–524. https://doi.org/10.1068/p150515

Callahan-Flintoft, C., Holcombe, A. O., & Wyble, B. (2019). A delay in sampling information from temporally autocorrelated visual stimuli. *BioRxiv*, 656850.

https://doi.org/10.1101/656850

Cameron, E. L., Eckstein, M., Tai, J., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks—Identification, yes/no detection and localization. *Spatial Vision*, *17*(4), 295–325. https://doi.org/10.1163/1568568041920212

Carlson, T. A., Hogendoorn, H., & Verstraten, F. A. J. (2006). The speed of visual attention: What time is it? *Journal of Vision*, *6*(12), 6–6. https://doi.org/10.1167/6.12.6

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Carrasco, M., Giordano, A. M., & McElree, B. (2006). Attention speeds processing across eccentricity: Feature and conjunction searches. *Vision Research*, *46*(13), 2028–2040. https://doi.org/10.1016/j.visres.2005.12.015

Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences*, *98*(9), 5363–5367.

Carrasco, M., McElree, B., Denisova, K., & Giordano, A. M. (2003). Speed of visual processing increases with eccentricity. *Nature Neuroscience*, *6*(7), 699.

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vision Research*, *40*(10–12), 1203–1215.

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *Journal of Vision*, *2*(6), 4–4.

Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 673.

Castiello, U., & Umiltà, C. (1990). Size of the attentional focus and efficiency of processing.

Acta Psychologica, *73*(3), 195–209. https://doi.org/10.1016/0001-6918(90)90022-8

Castiello, U., & Umiltà, C. (1992). Splitting focal attention. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 837.

Chakravarthi, R., & Cavanagh, P. (2007). Temporal properties of the polarity advantage effect in crowding. *Journal of Vision*, *7*(2), 11–11. https://doi.org/10.1167/7.2.11

Cheal, M., & Lyon, D. (1989). Attention Effects on Form Discrimination at Different Eccentricities. *The Quarterly Journal of Experimental Psychology Section A*, *41*(4), 719–746. https://doi.org/10.1080/14640748908402391

Cheal, M., & Lyon, D. R. (1991). Central and peripheral precuing of forced-choice discrimination. *The Quarterly Journal of Experimental Psychology*, *43*(4), 859–880.

Cheal, M., Lyon, D. R., & Hubbard, D. C. (1991). Does attention have different effects on line orientation and line arrangement discrimination? *The Quarterly Journal of Experimental Psychology*, *43*(4), 825–857.

Chun, M. M. (1997). Temporal binding errors are redistributed by the attentional blink. *Perception & Psychophysics*, *59*(8), 1191–1199. https://doi.org/10.3758/BF03214207

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*(1), 109.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, *27*(3), 183–228. https://doi.org/10.3758/BF03204258

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, *292*(4), 497–523. https://doi.org/10.1002/cne.902920402

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology*, *159*(2), 203–221.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual*

Review of Neuroscience, *18*(1), 193–222.

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

Dosher, B. A., & Lu, Z.-L. (2000a). Noise exclusion in spatial attention. *Psychological Science*, *11*(2), 139–146.

Dosher, B. A., & Lu, Z.-L. (2000b). Mechanisms of perceptual attention in precuing of location. *Vision Research*, *40*(10), 1269–1292. https://doi.org/10.1016/S0042-6989(00)00019-5

Du, F., & Abrams, R. A. (2010). Endogenous orienting is reduced during the attentional blink. *Experimental Brain Research*, *205*(1), 115–121. https://doi.org/10.1007/s00221-010-2345-9

Duncan, J. (1993). Similarity between concurrent visual discriminations: Dimensions and objects. *Perception & Psychophysics*, *54*(4), 425–430.

Dux, P. E., & Marois, R. (2009). The attentional blink: A review of data and theory. *Attention, Perception, & Psychophysics*, *71*(8), 1683–1700.

Eckstein, M. P. (1998). The Lower Visual Search Efficiency for Conjunctions Is Due to Noise and not Serial Attentional Processing. *Psychological Science*, *9*(2), 8.

Enns, J. T., & Di Lollo, V. (1997). Object Substitution: A New Form of Masking in Unattended Visual Locations. *Psychological Science*, *8*(2), 135–139. https://doi.org/10.1111/j.1467-9280.1997.tb00696.x

Enns, J. T., & Di Lollo, V. (2000). What’s new in visual masking? *Trends in Cognitive Sciences*, *4*(9), 345–352. https://doi.org/10.1016/S1364-6613(00)01520-5

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*(2), 201–204.

Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, *14*(1), 155–160.

Eriksen, C. W., & Yeh, Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 583.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303. https://doi.org/10.1038/nrn2258

Foley, J. M., & Schwarz, W. (1998). Spatial attention: Effect of position uncertainty and number of distractor patterns on the threshold-versus-contrast function for contrast discrimination. *Journal of the Optical Society of America A*, *15*(5), 1036. https://doi.org/10.1364/JOSAA.15.001036

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030.

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, *8*(4), 215–221.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. https://doi.org/10.1038/nn.2889

Gathercole, S. E., & Broadbent, D. E. (1984). Combining attributes in specified and categorized target search: Further evidence for strategic differences. *Memory & Cognition*, *12*(4), 329–337. https://doi.org/10.3758/BF03198292

Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time*. Department of Statistics, Columbia University. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, *60*(4), 328–331.

Giordano, A. M., McElree, B., & Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *Journal of Vision*, *9*(3), 30–30. https://doi.org/10.1167/9.3.30

Goodbourn, P. T., & Holcombe, A. O. (2015). “Pseudoextinction”: Asymmetries in simultaneous attentional selection. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 364.

Goodbourn, P. T., Martini, P., Barnett-Cowan, M., Harris, I. M., Livesey, E. J., & Holcombe, A. O. (2016). Reconsidering Temporal Selection in the Attentional Blink. *Psychological Science*, *27*(8), 1146–1156. https://doi.org/10.1177/0956797616654131

Grossberg, S., & Stone, G. (1986). Neural dynamics of attention switching and temporal-order information in short-term memory. *Memory & Cognition*, *14*(6), 451–468. https://doi.org/10.3758/BF03202517

Hamilton, R. H., Stark, M., & Coslett, H. B. (2010). Increased effect of target eccentricity on covert shifts of visual attention in patients with neglect. *Cortex*, *46*(1), 68–76. https://doi.org/10.1016/j.cortex.2009.02.005

Harrison, W. J., & Bex, P. J. (2015). A Unifying Model of Orientation Crowding in Peripheral Vision. *Current Biology*, *25*(24), 3213–3219. https://doi.org/10.1016/j.cub.2015.10.052

Hartmann, E., Lachenmayr, B., & Brettel, H. (1979). The peripheral critical flicker frequency. *Vision Research*, *19*(9), 1019–1023. https://doi.org/10.1016/0042-6989(79)90227-X

Hawkins, R. X. D., Houpt, J. W., Eidels, A., & Townsend, J. T. (2016). Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. *Vision Research*, *126*, 19–33. https://doi.org/10.1016/j.visres.2015.04.019

Hein, E., Rolke, B., & Ulrich, R. (2006). Visual attention and temporal discrimination:

Differential effects of automatic and voluntary cueing. *Visual Cognition*, *13*(1), 29–50.

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2019). *A conceptual introduction to Bayesian Model Averaging* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/wgb64

Hogendoorn, H., Carlson, T. A., Vanrullen, R., & Verstraten, F. A. (2010). Timing divided attention. *Attention, Perception, & Psychophysics*, *72*(8), 2059–2068.

Holcombe, A. O., & Cavanagh, P. (2008). Independent, synchronous access to color and motion features. *Cognition*, *107*(2), 552–580. https://doi.org/10.1016/j.cognition.2007.11.006

Holcombe, A. O., Nguyen, E. H., & Goodbourn, P. T. (2017). Implied reading direction and prioritization of letter encoding. *Journal of Experimental Psychology: General*, *146*(10), 1420.

Huang, L., & Dobkins, K. R. (2005). Attentional effects on contrast discrimination in humans: Evidence for both contrast gain and response gain. *Vision Research*, *45*(9), 1201–1212. https://doi.org/10.1016/j.visres.2004.10.024

Huckauf, A., & Heller, D. (2004). On the relations between crowding and visual masking. *Perception & Psychophysics*, *66*(4), 584–595. https://doi.org/10.3758/BF03194903

Jefferies, L. N., Ghorashi, S., Kawahara, J.-I., & Di Lollo, V. (2007). Ignorance is bliss: The role of observer expectation in dynamic spatial tuning of the attentional focus. *Perception & Psychophysics*, *69*(7), 1162–1174. https://doi.org/10.3758/BF03193953

Jeffreys, H. (1998). *The theory of probability* (3rd ed.). OUP Oxford.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, *43*(4), 346–354.

Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, 404–425.

Kastner, S., de Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of

Directed Attention in the Human Extrastriate Cortex as Revealed by Functional MRI. *Science*, *282*(5386), 108–111. JSTOR.

Kastner, S., Weerd, P. D., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *J. Neurophysiol*, 1398–1411.

Kehrer, L. (1989). Central performance drop on perceptual segregation tasks. *Spatial Vision*, *4*(1), 45–62. https://doi.org/10.1163/156856889X00040

Kennedy, G. J., & Whitaker, D. (2010). The chromatic selectivity of visual crowding. *Journal of Vision*, *10*(6), 15–15. https://doi.org/10.1167/10.6.15

King, J.-R., & Wyart, V. (2019). The Human Brain encodes a Chronicle of Visual Events at each Instant of Time. *BioRxiv*, 846576. https://doi.org/10.1101/846576

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, *8*(2), 255–279.

Kristjánsson, Á., & Nakayama, K. (2002). The attentional blink in space and time. *Vision Research*, *42*(17), 2039–2050.

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: Attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641), 20130205. https://doi.org/10.1098/rstb.2013.0205

Lawrence, D. H. (1971). Two studies of visual search for word targets with controlled rates of presentation. *Perception & Psychophysics*, *10*(2), 85–89.

Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *Journal of the Optical Society of America A*, *22*(10), 2013. https://doi.org/10.1364/JOSAA.22.002013

Levi, D. M., Hariharan, S., & Klein, S. A. (2002a). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision*, *2*(2), 3–3. https://doi.org/10.1167/2.2.3

Levi, D. M., Hariharan, S., & Klein, S. A. (2002b). Suppressive and facilitatory spatial interactions in amblyopic vision. *Vision Research*, *42*(11), 1379–1394. https://doi.org/10.1016/S0042-6989(02)00061-5

Ling, S., & Carrasco, M. (2006). Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research*, *46*(8–9), 1210–1220. https://doi.org/10.1016/j.visres.2005.05.008

Liu, T., Pestilli, F., & Carrasco, M. (2005). Transient Attention Enhances Perceptual Performance and fMRI Response in Human Visual Cortex. *Neuron*, *45*(3), 469–477. https://doi.org/10.1016/j.neuron.2004.12.039

Liu, T., Stevens, S. T., & Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research*, *47*(1), 108–113.

Low, F. N. (1951). Peripheral visual acuity. *Archives of Ophthalmology*, *45*(1), 80–99.

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*(9), 1183–1198.

Lu, Z.-L., & Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A*, *16*(3), 764. https://doi.org/10.1364/JOSAA.16.000764

Lu, Z.-L., & Dosher, B. A. (2000). Spatial attention: Different mechanisms for central and peripheral temporal precues? *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1534.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616.

Lunau, R., & Olivers, C. N. L. (2010). The attentional blink and lag 1 sparing are nonspatial. *Attention, Perception, & Psychophysics*, *72*(2), 317–325.

Madden, D. J. (1992). Selective attention and visual search: Revision of an allocation model and application to age differences. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 821.

Maki, W. S., Frigen, K., & Paulson, K. (1997). Associative priming by targets and distractors during rapid serial visual presentation: Does word meaning survive the attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 1014–1034. http://dx.doi.org/10.1037/0096-1523.23.4.1014

Martens, S., & Wyble, B. (2010). The attentional blink: Past, present, and future of a blind spot in perceptual awareness. *Neuroscience & Biobehavioral Reviews*, *34*(6), 947–957.

Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nature Communications*, *8*(1), 1955. https://doi.org/10.1038/s41467-017-02079-x

McElree, B., & Carrasco, M. (1999). The Temporal Dynamics of Visual Search: Evidence for Parallel Processing in Feature and Conjunction Searches. *Journal of Experimental Psychology*, *25*(6), 1517–1539.

McLean, J. P., Broadbent, D. E., & Broadbent, M. H. (1983). Combining attributes in rapid serial visual presentation tasks. *The Quarterly Journal of Experimental Psychology*, *35*(1), 171–186.

Montagna, B., Pestilli, F., & Carrasco, M. (2009). Attention trades off spatial acuity. *Vision Research*, *49*(7), 735–745. https://doi.org/10.1016/j.visres.2009.02.001

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs*. https://richarddmorey.github.io/BayesFactor/

Moutoussis, K., & Zeki, S. (1997). A direct demonstration of perceptual asynchrony in vision. *Proceedings of the Royal Society of London B: Biological Sciences*, *264*(1380), 393–399.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 315.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*(11), 1631–1647.

Ng, J., & Westheimer, G. (2002). Time Course of Masking in Spatial Resolution Tasks: *Optometry and Vision Science*, *79*(2), 98–102. https://doi.org/10.1097/00006324-200202000-00011

Nieuwenstein, M. R., & Potter, M. C. (2006). Temporal Limits of Selection and Memory Encoding: A Comparison of Whole Versus Partial Report in Rapid Serial Visual Presentation. *Psychological Science*, *17*(6), 471–475. https://doi.org/10.1111/j.1467-9280.2006.01730.x

Nobre, A. C., Correa, A., & Coull, J. T. (2007). The hazards of time. *Current Opinion in Neurobiology*, *17*(4), 465–470.

Nordfang, M., & Wolfe, J. M. (2014). Guided search for triple conjunctions. *Attention, Perception, & Psychophysics*, *76*(6), 1535–1559. https://doi.org/10.3758/s13414-014-0715-2

Olivers, C. N. L., & Meeter, M. (2008). A boost and bounce theory of temporal attention. *Psychological Review*, *115*(4), 836.

Olivers, C. N. L., van der Stigchel, S., & Hulleman, J. (2007). Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research*, *71*(2), 126–139. https://doi.org/10.1007/s00426-005-0029-z

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*(5), 1. https://doi.org/10.1167/5.5.1

Pelli, D. G., & Blakemore, C. (1990). The quantum efficiency of vision. *Vision: Coding and*

Efficiency, 3–24.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*(28), 4646–4674. https://doi.org/10.1016/j.visres.2006.04.023

Pelli, D. G., & Farell, B. (1999). Why use noise? *JOSA A*, *16*(3), 647–653.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12), 12–12. https://doi.org/10.1167/4.12.12

Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences*, 187–199.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, *11*(10), 1129–1135. https://doi.org/10.1038/nn.2187

Pestilli, F., & Carrasco, M. (2005). Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research*, *45*(14), 1867–1875. https://doi.org/10.1016/j.visres.2005.01.019

Pinto, Y., Sligte, I. G., Shapiro, K. L., & Lamme, V. A. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin & Review*, *20*(4), 732–739.

Poggel, D. A., & Strasburger, H. (2004). Visual perception in space and time-mapping the visual field of temporal resolution. *Acta Neurobiologiae Experimentalis*, *64*(3), 427–437.

Poggel, D. A., Treutwein, B., Calmanti, C., & Strasburger, H. (2006). Increasing the temporal g(r)ain: Double-pulse resolution is affected by the size of the attention focus. *Vision Research*, *46*(18), 2998–3008. https://doi.org/10.1016/j.visres.2006.03.018

Popple, A. V., & Levi, D. M. (2007). Attentional blinks as errors in temporal binding. *Vision Research*, *47*(23), 2973–2981. https://doi.org/10.1016/j.visres.2007.06.022

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, *32*, 531–556.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509.

Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, *21*(2), 156–161.

Potter, M. C., Kroll, J. F., & Harris, C. (1981). Comprehension and Memory in Rapid Sequential Reading. In R. S. Nickerson (Ed.), *Attention and Performance: Vol. VIII* (p. 24). Erlbaum.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*(1), 10.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raftery, A. E. (1999). Bayes factors and BIC: Comment on “A critique of the Bayesian information criterion for model selection.” *Sociological Methods & Research*, *27*(3), 411–427.

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*(3), 849.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review*, *93*(2), 180.

Remington, R. W., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Attention, Perception, & Psychophysics*, *51*(3), 279–290.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*(5), 1736–1753.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17–17. https://doi.org/10.1167/7.2.17

Rosenholtz, R., Yu, D., & Keshvari, S. (2019). Challenges to pooling models of crowding: Implications for visual mechanisms. *Journal of Vision*, *19*(7), 15–15. https://doi.org/10.1167/19.7.15

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Rusconi, P., & Huber, D. E. (2018). The perceptual wink model of non-switching attentional blink tasks. *Psychonomic Bulletin & Review*, *25*(5), 1717–1739. https://doi.org/10.3758/s13423-017-1385-6

Saslow, M. G. (1967). Effects of Components of Displacement-Step Stimuli Upon Latency for Saccadic Eye Movement. *Journal of the Optical Society of America*, *57*(8), 1024. https://doi.org/10.1364/JOSA.57.001024

Scalf, P. E., & Beck, D. M. (2010). Competition in Visual Cortex Impedes Attention to Multiple Items. *Journal of Neuroscience*, *30*(1), 161–169. https://doi.org/10.1523/JNEUROSCI.4207-09.2010

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*(2), 95–100.

Shih, S.-I. (2000). Recall of two visual targets embedded in RSVP streams of distractors depends on their temporal and spatial relationship. *Perception & Psychophysics*, *62*(7), 1348–1355. https://doi.org/10.3758/BF03212137

Shih, S.-I., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review*, *109*(2), 260.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are There Multiple Visual Short-Term Memory Stores? *PLOS ONE*, *3*(2), e1699. https://doi.org/10.1371/journal.pone.0001699

Smith, W. S., Mollon, J. D., Bhardwaj, R., & Smithson, H. E. (2011). Is there brief temporal buffering of successive visual inputs? *The Quarterly Journal of Experimental Psychology*, *64*(4), 767–791. https://doi.org/10.1080/17470218.2010.511237

Smithson, H., & Mollon, J. (2006). Do masks terminate the icon? *Quarterly Journal of Experimental Psychology*, *59*(1), 150–160. https://doi.org/10.1080/17470210500269345

Solomon, J. A., Lavie, N., & Morgan, M. J. (1997). Contrast discrimination function: Spatial cuing effects. *Journal of the Optical Society of America A*, *14*(9), 2443. https://doi.org/10.1364/JOSAA.14.002443

Spencer, T. J., & Shuntich, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology*, *85*(2), 198–203.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1.

Strasburger, H. (2005). Unfocussed spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, *5*(11), 8–8. https://doi.org/10.1167/5.11.8

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, *49*(6), 495–508.

Strasburger, H., Rentschler, I., & Harvey, L. O. (1994). Cortical Magnification Theory Fails to Predict Visual Recognition. *European Journal of Neuroscience*, *6*(10), 1583–1588. https://doi.org/10.1111/j.1460-9568.1994.tb00548.x

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13–13. https://doi.org/10.1167/11.5.13

Suantak, L., Bolger, F., & Ferrell, W. R. (1996). The Hard–Easy Effect in Subjective Probability Calibration. *Organizational Behavior and Human Decision Processes*, *67*(2), 201–221. https://doi.org/10.1006/obhd.1996.0074

Swets, J. A., Tanner, W. P., Jr, & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*(5), 301.

Thom, H. C. (1958). A note on the gamma distribution. *Monthly Weather Review*, *86*(4), 117–122.

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

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141. https://doi.org/10.1016/0010-0285(82)90006-8

Tyler, C. W. (1987). Analysis of visual modulation sensitivity III Meridional variations in peripheral flicker sensitivity. *Journal of the Optical Society of America A*, *4*(8), 1612. https://doi.org/10.1364/JOSAA.4.001612

Verghese, P. (2001). Visual Search and Attention: A Signal Detection Theory Approach. *Neuron*, *31*(4), 523–535. https://doi.org/10.1016/S0896-6273(01)00392-0

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*(6), 1656–1674. https://doi.org/10.1037/0096-1523.24.6.1656

Vossel, S., Thiel, C. M., & Fink, G. R. (2006). Cue validity modulates the neural correlates of covert endogenous orienting of attention in parietal and frontal cortex. *NeuroImage*, *32*(3), 1257–1264. https://doi.org/10.1016/j.neuroimage.2006.05.019

Vul, E., Hanus, D., & Kanwisher, N. (2008). Delay of selective attention during the attentional blink. *Vision Research*, *48*(18), 1902–1909. https://doi.org/10.1016/j.visres.2008.06.009

Vul, E., Nieuwenstein, M., & Kanwisher, N. (2008). Temporal selection is suppressed, delayed, and diffused during the attentional blink. *Psychological Science*, *19*(1), 55–61.

Vul, E., & Rich, A. N. (2010). Independent Sampling of Features Enables Conscious Perception of Bound Objects. *Psychological Science*, *21*(8), 1168–1175. https://doi.org/10.1177/0956797610377341

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Weichselgartner, E., & Sperling, G. (1987). Dynamics of Automatic and Controlled Visual Attention. *Science*, *238*(4828), 778–780.

Westheimer, G. (2004). Center-surround antagonism in spatial vision: Retinal or cortical locus? *Vision Research*, *44*(21), 2457–2465. https://doi.org/10.1016/j.visres.2004.05.014

Weymouth, F. W. (1958). Visual Sensory Units and the Minimal Angle of Resolution*. *American Journal of Ophthalmology*, *46*(1), 102–113. https://doi.org/10.1016/0002-9394(58)90042-4

White, J. A., Rubinstein, J. T., & Kay, A. R. (2000). Channel noise in neurons. *Trends in Neurosciences*, *23*(3), 131–137. https://doi.org/10.1016/S0166-2236(99)01521-0

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005

Wolfe, J. M. (1998). What Can 1 Million Trials Tell Us About Visual Search? *Psychological*

Science, *9*(1), 33–39. https://doi.org/10.1111/1467-9280.00006

Wolfe, J. M., & Gray, W. (2007). Guided search 4.0. *Integrated Models of Cognitive Systems*, 99–119.

Wundt, W. (1897). *Outline of psychology.* (C. H. Judd, Trans.). Wilhelm Engelmann. https://doi.org/10.1037/12908-000

Wyble, B., Bowman, H., & Nieuwenstein, M. (2009). The attentional blink provides episodic distinctiveness: Sparing at a cost. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 787–807. https://doi.org/10.1037/a0013902

Yantis, S. (1998). Control of visual attention. In H. E. Pashler (Ed.), *Attention* (pp. 223–256). Psychology Press/Erlbaum (UK) Taylor & Francis.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 601.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*(6706), 72. https://doi.org/10.1038/23936

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, *39*(2), 293–306.

Yeshurun, Y., Rashal, E., & Tkacz-Domb, S. (2015). Temporal crowding and its interplay with spatial crowding. *Journal of Vision*, *15*(3), 11–11. https://doi.org/10.1167/15.3.11

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860