THEORETICAL/REVIEW

The diachronic account of attentional selectivity

Alon Zivony[1] · Martin Eimer[1]

Abstract

Many models of attention assume that attentional selection takes place at a specific moment in time that demarcates the critical transition from pre-attentive to attentive processing of sensory input. We argue that this intuitively appealing standard account of attentional selectivity is not only inaccurate, but has led to substantial conceptual confusion. As an alternative, we offer a 'diachronic' framework that describes attentional selectivity as a process that unfolds over time. Key to this view is the concept of attentional episodes, brief periods of intense attentional amplification of sensory representations that regulate access to working memory and response-related processes. We describe how attentional episodes are linked to earlier attentional mechanisms and to recurrent processing at the neural level. We review studies that establish the existence of attentional episodes, delineate the factors that determine if and when they are triggered, and discuss the costs associated with processing multiple events within a single episode. Finally, we argue that this framework offers new solutions to old problems in attention research that have never been resolved. It can provide a unified and conceptually coherent account of the network of cognitive and neural processes that produce the goal-directed selectivity in perceptual processing that is commonly referred to as 'attention'.

Keywords Attentional selection · Recurrent processing · Attentional episodes · Attention systems

Introduction

After more than a century of research, attention researchers have come full circle from the intuition that "everybody knows what attention is" (James, 1890) to the conclusion that "nobody knows what attention is" (Hommel et al., 2019). Indeed, some researchers hold to the pessimistic view that "attention is a sterile concept" (Di Lollo, 2018), and suggest that this concept is incoherent and unsuitable for scientific research (e.g., Anderson, 2011; Allport, 1993), and should therefore best be abandoned.

We propose a more optimistic view. Referring to selective attention can play an important role in explaining perception and performance, as long as this term is clearly defined, and is linked both to underlying neural mechanisms and a wide range of cognitive phenomena. We take a step towards these goals by first providing a diagnosis of why current concepts

of selective attention are problematic, and then developing an alternative framework. We provide the outlines of a new model of attentional selectivity in visual processing that focuses specifically on its temporal aspects, and highlights the fact that the effects of selective attention at the cognitive and neural level emerge gradually and change across time. Our emphasis on the time course of attentional selectivity is by no means new. Almost a century ago, Titchener (1928) discussed whether a "single wave of attention" can be internally maintained for extended periods, anticipating what has since become known as research on sustained attention. In contrast to Titchener, we focus on attentional mechanisms that operate within a shorter time frame, in the first few hundred milliseconds after potentially relevant visual stimuli have been presented. Most importantly, we build on previous models that describe attentional selectivity as the result of mechanisms that unfold continuously in real time (e.g., Olivers & Meeter, 2008; Reeves & Sperling, 1986; Wyble et al., 2011). In this framework, attentional selection is defined not as single discrete operation or stage, but instead as a temporally extended process. For this reason, we refer to our and other related accounts as 'diachronic' models of selective

✉ Alon Zivony
  alonzivony@gmail.com

[1] Department of Psychological Sciences, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK

attention.[1] An important part of these models is the assumption that the core function of selective attention –allowing objects that are relevant in the context of current task goals access to working memory and action control – takes place within an extended temporal window ('attentional episode').

We begin with a brief historical review of standard (non-diachronic) accounts of attention, and discuss why these accounts have contributed to decades of unresolved debate and conceptual confusion. We then present the outlines of an alternative diachronic framework, building on previous diachronic models to introduce a unified diachronic account of attentional selectivity (UDAS) and its three main premises. The account integrates key insights from previous diachronic models into a general theoretical and non-computational account, and connects them to the broader psychological and neuroscientific literature on selective attention. We discuss the roles of spatial, feature-based, and temporal attention for the control of attentional episodes, and provide a brief outline of the neural basis of attentional selectivity within this diachronic framework. We then evaluate current empirical evidence for the existence of attentional episodes, the costs that arise when multiple objects are processed within the same episode, and consider the adaptive functions of attentional episodes in visual cognition. Finally, we take a look at the past and future of attention research, and argue that a new focus on the diachronic nature of attentional selectivity can help to resolve many of the problems that still block progress in our field. In this review, we focus primarily on cases where attention is guided in the absence of overt eye movements to specific task-relevant objects (covert endogenous attention), as in tasks where these objects appear in visual search displays, rapid serial visual presentation (RSVP) streams, or follow a spatially predictive cue. Other aspects of attentional selectivity, such as the guidance of eye movements (overt attention shifts) and the impact of selection history, are only briefly discussed.

The standard non-diachronic account of attentional selection

Many of the core assumptions in current attention research can be traced back to Broadbent's (1958) seminal filter model. This model postulates an essentially serial feedforward architecture for perception and recognition. Incoming sensory information is processed along a unidirectional pathway of discrete stages, eventually resulting in some of it being encoded in memory repositories and becoming available to deliberate response-related processes. Another assumption is that early sensory stages are automatic and can process virtually unlimited amount of information, whereas subsequent stages are severely capacity-limited. To prevent overload, access to these later stages is regulated by an attentional filtering mechanism that determines which sensory signals receives additional processing and which are discarded. This attentional selection process is at the core of Broadbent's model. It conceptualizes attention as a causal agent (Johnston & Dark, 1986) that determines the fate of incoming perceptual information, and entails a neat and fundamental dichotomy between pre-attentive and attentive processes (Neisser, 1967), with 'selection' at the interface between them.

This framework has inspired decades of attention research (see Allport, 1993, for a review) that attempted to dissociate pre-attentive and attentive processes, and to identify the locus of attentional selection (i.e., where in the linear feedforward processing architecture this transition occurs). However, since its inception, core assumptions of the filter model have been thoroughly debated and criticised (e.g., Allport, 1993; Driver, 2001; Johnston & Dark, 1986). Influenced by developments in neuroscience, it is now widely agreed that perceptual mechanisms do not exclusively operate in a serial linear fashion, but also involve extensive parallel processing (e.g., Desimone & Duncan, 1995). In addition, feedforward perceptual pathways are generally accompanied by various feedback loops that enable iterative recurrent processing (e.g., Lamme & Roelfsema, 2000). Moreover, attention is no longer seen as a unitary mechanism serving a specific distinct function, but is now understood as a network of multiple interrelated processes supported by different brain regions, serving separable functions such as alerting, orienting, and executive control (e.g., Petersen & Posner, 2012). Selective attention can also operate in different domains, facilitating the processing of specific locations (spatial attention), specific features (feature-based attention), and specific points in time (temporal attention). These insights have resulted in a fragmentation of attention research into multiple sub-fields, and have made it even harder to conceptualize attention as a single coherent concept (Hommel et al., 2019).

However, another core assumption in Broadbent's model has so far largely escaped critical scrutiny and is still widely embraced in current attention research, albeit often implicitly: the attentional filtering process takes place at a specific moment in time. This assumption logically follows from Broadbent's linear architecture, and his central suggestion that selection demarcates the point in a serial feedforward chain where sensory information is

---

[1] The Merriam-Webster dictionary defines 'diachronic' – stemming from the Greek words for 'through' (δια) and 'time' (χρόνος) – as "relating to, or dealing with, phenomena as they occur or change over a period of time". This term is commonly used in linguistics to refer to the study of language or cultural developments that often occur over the span of decades and centuries. Here, we use this term to describe the cognitive and neural architecture responsible for the selective processing of visual information that operates within a much shorter time frame of a few hundreds of milliseconds.

transferred from a parallel capacity-unlimited stage to a subsequent serial capacity-limited stage. Selection is thus understood as an all-or-none stage that occurs at a specific point in time within this serial feedforward architecture. Once it has taken place, an event either gains privileged access to post-selection processes such as working memory encoding and response selection, or is discarded from further processing.

Broadbent's idea of a temporally discrete all-or-none filter mechanism remains influential in many contemporary accounts of attentional selectivity. Its tell-tale sign is the assumption that there is a specific point in time when particular events are selected for access to higher-level cognitive process, while others are ignored. Selection may be based on bottom-up saliency (Theeuwes, 2010), top-down goals (Folk et al., 1992; Folk & Remington, 1998), selection history (Failing & Theeuwes, 2018), or reward value (Anderson, 2016), and may occur early during perceptual processing (Treisman, 1998) or only after perceptual analysis is completed (Chun & Potter, 1995). In these accounts, a discrete concept of 'attentional selection' is retained in much the same way as it was originally conceived in Broadbent's filter model, in spite of the fact that the serial feedforward architecture on which this concept was based is no longer regarded as a realistic description of perceptual processing. For example, a highly influential model of visual search (Guided Search; e.g., Wolfe, 2021) describes how pre-attentive guidance processes (combining bottom-up and top-down signals within a single priority map) feed into a temporally discrete selection stage. This stage constitutes a serial bottleneck where individual objects are selected to gain access to subsequent post-attentive recognition processes. While both guidance and recognition are temporally extended and can process multiple objects in parallel, Guided Search retains Broadbent's core idea that these pre- and post-selection processes are separated by a discrete serial selection stage that operates at a specific moment in time. More generally, even though the existence of multiple attentional mechanisms and recurrent processing is now widely acknowledged, the idea that there is a fundamental distinction between pre-attentive and attentive processes (with 'selection' marking the discrete interface between these processes) remains popular. As a result, many recent studies still investigate whether specific processes fall on one or the other side of this divide (e.g., Drisdelle & Jolicoeur, 2018; Itthipuripat et al., 2019; Perrone-Bertolotti et al., 2020; Tamber-Rosenau & Marois, 2016).

We suggest that this concept of selection as a distinct step in the processing of visual information that operates at a particular moment in time is responsible for much of the incoherence in how attention is conceptualized. For example, the concept of attentional selection as a temporally discrete mechanism retains the ambiguity about whether attention should be thought of as a causal agent or an effect (Johnston & Dark, 1986). Selection is often described as the mechanism that distinguishes between relevant and irrelevant information, without clear specification of the basis of this ability. Apart from the obvious danger of postulating an intelligent homunculus, such ambiguity in the way that the concept of selection is employed can also result in circularity: if a specific manipulation results in the encoding of an event in memory, attentional selection is both inferred from the result and assumed to be its cause (Di Lollo, 2018). In cases like this, attentional selection is not a useful explanatory construct but rather an unnecessary descriptor that might just as well be omitted (Anderson, 2011). Worse still, retaining a discrete conception of attentional selectivity can result in the misleading description of continuous and dynamic processes such as attentional capture as temporally distinct ('quantile') events (Anderson, 2021), which might impede scientific progress in the field.

The diachronic account of attention and the importance of attentional episodes

How can we account for the evident selective nature of perceptual processing while avoiding references to a problematic discrete notion of selection? The key insight is that attentional processes unfold gradually across time, and as such, result in varying degrees of selectivity at any given moment. Here, we propose to abandon the standard concept of selection and replace it with a *diachronic* account of selectivity that emphasizes the critical temporal aspects of attentional processing.

From a neurophysiological perspective, the fact that attentional selectivity operates in the time domain is obvious. Attentional modulations of neural activity emerge at specific moments in time, and are either transient (Hayden & Gallant, 2005) or maintained over longer periods, such as the sustained baseline shifts observed during attentional preparation (e.g., Driver & Frith, 2000; Kanwisher & Wojciulik, 2000). These temporal aspects of attention are often ignored in cognitive theories of selective attention, even those that attempt to connect attention with neural mechanisms. Consider, for example, Treisman's suggestion that focused spatial attention facilitates the binding of object features that are processed in different parts of the visual cortex. Her famous suggestion that attention provides the 'glue' for feature binding by enhancing the neural firing of retinotopically mapped visual neurons (Treisman, 1996, 1998) emphasizes space but does not account for binding in the temporal domain. In dynamic visual environments where objects appear, disappear, or move at rapid rates, feature binding must occur not only in space but also in time. This temporal binding problem is exacerbated by the fact that different features

are processed at different speeds (Treisman, 2014; Wolfe, 2014), and by the overlap of fast feedforward and subsequent recurrent processing in the same brain areas (e.g., Fahrenfort et al., 2008; Rademaker et al., 2019). To avoid feature misbinding under such conditions, attentional selectivity needs to include a temporal component that keeps track of the continuity of visual objects across time.

A small number of models of attention, which we refer to as *diachronic models*, focus on these temporal aspects. The first of these models was proposed by Reeves and Sperling (1986). Their attentional gating theory is a computational model that was developed to explain order reversals in visual reports (see below). At the heart of this model is the assumption that selectivity is implemented through brief periods of amplified perceptual processing at specific locations within retinotopic maps, which selectively increases the activation levels of particular sensory representations (i.e., representations of features of visual objects such as their colour, size, or shape, that are generated at relatively early stages of perceptual processing prior to the recognition of these objects and their classification into task-relevant categories). This attentional amplification of perceptual processing is not uniform, but is assumed to build up rapidly, reach a peak after 100–200 ms, and then gradually dissipates ("attentional episodes"). Similar ideas about temporal windows for selective attentional processing have also been proposed and developed in several other recent diachronic models (Bowman & Wyble, 2007; Olivers & Meeter, 2008; Põder, 2013; Shih, 2008; Shih & Reeves, 2007; Smith & Wolfgang, 2004; Wyble et al., 2009; Wyble et al., 2011).

While some of these ideas have been useful to account for phenomena such as the attentional blink (AB, see below for details), the general importance of attentional episodes for theoretical models of attentional selectivity has so far largely been overlooked. This is particularly the case for non-computational models that aim to explain attentional effects that are averaged across multiple trials. In contrast to computational models that simulate such effects on a single-trial level, averaged data can readily be explained in terms of discrete selection processes that are temporally variable across trials (see Vul et al., 2009, for discussion). Here, we develop a unified diachronic account of attentional selectivity (UDAS) that draws on previous ideas developed by Reeves and Sperling, Wyble, and others. While we outline this new framework in a general non-formalized way, many aspects of this account should lend themselves readily to formal computational modelling. Following Wyble et al. (2009, 2011), we use the term *attentional episode* to refer to the entire period of perceptual amplification, and *attentional engagement* to refer to the start of the attentional episode (see Zivony & Eimer, 2021a, for further details). This model explicitly rejects the existence of a temporally discrete all-or-none attentional selection stage, and the

resulting fundamental division between parallel pre-attentive and serial attentive processes. Instead, the UDAS describes selectivity as a temporally extended process. This process is initiated and controlled by goal-sensitive mechanisms that operate at early parallel sensory-perceptual stages, followed by a critical phase of spatially specific enhancement of visual representations (attentional episode), which modulate the ability of these representations to gain access to working memory (WM), response selection, and conscious awareness.

Three premises of the UDAS

Premise 1: Evidence accumulation

Similar to other models (e.g., Reeves & Sperling, 1986), the UDAS describes perception as an evidence accumulation process that extracts information about visual objects (e.g., their features or identities) by generating sensory representations of these objects. This perceptual sampling process operates in parallel across the entire visual field, is inherently noisy (Ashby & Lee, 1993), and fluctuates in its efficiency across different spatial locations and across time. As a result, the quality (or 'strength') of specific sensory representations is also variable (see Fig. 1c and e). These representations tend to dissipate rapidly after stimulus offset, and their quality can also be reduced by inhibitory interactions with temporally and/or spatially adjacent stimuli (perceptual competition; Wyble et al., 2011). Evidence accumulation is not contingent on attentional amplification (i.e., sampling occurs also for unattended objects). However, a central assumption of the UDAS is that attentional episodes strongly modulate evidence accumulation for particular objects and events. They do so by substantially speeding the rate of sampling from a specific location during the amplification period, which modulates the strength of perceptual representations at that location. The strength of sensory representations affects their durability across time and is also a crucial factor in determining whether they gain access to stages of processing such as WM encoding and response selection.

Premise 2: Engagement threshold

Attentional episodes are triggered by external events, but not every stimulus necessarily engages attention. The UDAS assumes that attentional episodes are triggered by the detection of a potential target object or target-defining features (Zivony & Lamy, 2018). The parameters relevant to distinguish between task-relevant and irrelevant sensory signals are represented by attentional templates, which are typically activated during the preparation for an attentional selection task (Grubert & Eimer, 2018), and are assumed to be stored in WM (Bundesen et al., 2005). These templates are matched
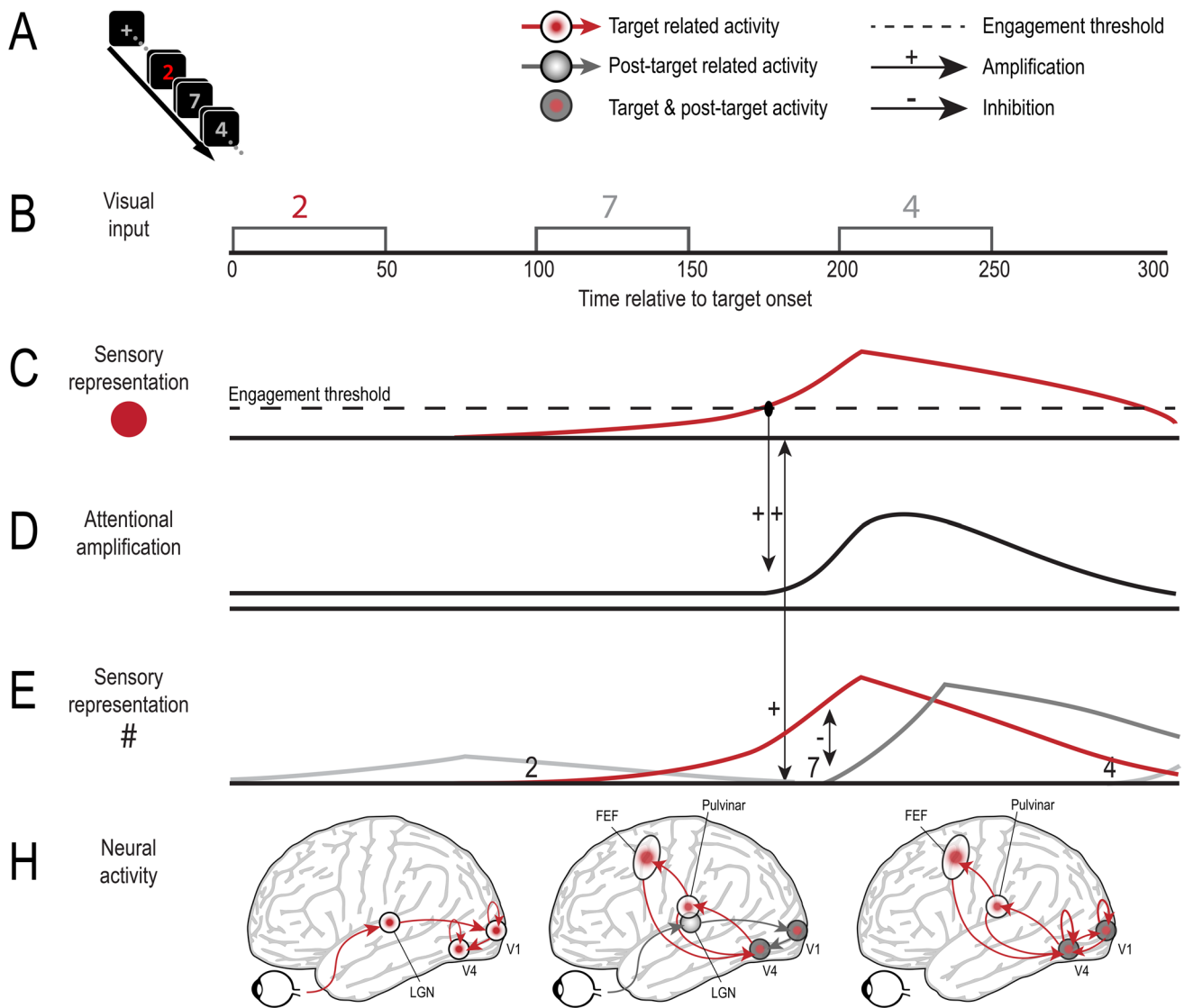
Fig. 1 Schematic outline of the time course of attentional modulations of early sensory processing and subsequent attentional episodes. In this example (**a**), the target is a red digit among grey digits presented in a rapid serial visual presentation (RSVP) task where stimuli are presented at rate of 10 Hz (**b**). Visual processing starts 50–80 ms after stimulus onset, resulting in weak evidence accumulation about the colour (**c**) and shape (**e**) of items in the RSVP stream. Approximately 170 ms after target onset, the activation of the sensory representation of the target-defining colour (**c**) reaches the engagement threshold (dashed line), and an attentional episode is triggered (**d**). During this episode, the processing of all sensory representations is amplified (**e**), including the representation of the post-target distractor (grey line). Representations within the episode are subject to competitive interactions. Bottom panel (**f**): Cortical basis of attentional modulations occurring prior to and during attentional episodes. Activation levels are modulated during the initial feedforward sweep of visual processing (left panel), indicated by local facilitation. Once activation reaches a critical threshold value, recurrent signals from the pulvinar and frontal eye field (FEF) to extrastriate visual cortex trigger an attentional episode (middle panel). These recurrent signals remain activated throughout the duration of the episode, amplifying both target processing as well as the processing of the subsequent post-target distractor (right panel)

to perceptual evidence that accumulates in parallel and the strength of this evidence has to cross a hypothetical engagement threshold (see Fig. 1c) before an attentional episode is triggered at a particular location. Due to the noisy nature of evidence accumulation, the point in time where this threshold is reached is inherently variable (Zivony & Eimer, 2021a).

Premise 3: Indiscriminate attentional amplification

Once the engagement threshold is crossed, an attentional episode is triggered and persists independently of the continued presence of the triggering event. The duration of the attentional episode may be affected by the nature of the item(s)

that follow a task-relevant event at the same location (Olivers & Meeter, 2008; Reeves & Sperling, 1986; Wyble et al., 2011), but typically lasts for approximately 150 ms. All items that appear at this attended location during the attentional episode will benefit from amplification (see Fig. 1d), making them more durable and more likely to be encoded in WM, irrespective of whether they are task-relevant or not. This indiscriminate attentional amplification principle can account for various phenomena that have been studied by cognitive scientists in the past 40 years (see below).

This description of an attentional episode retains a discrete element, as amplification is triggered at a specific moment in time, following above-threshold activation of target-defining feature representations. However, and critically, selectivity in this diachronic framework is not understood as a discrete time point where 'selection' takes place. Instead, it is conceptualized as a gradually emerging process that includes changes in the activation levels of feature representations that occur before, during, and after an attentional episode. Figure 1 illustrates the implications of the three premises of the UDAS for the time course of attentional selectivity in the processing of successive items presented in an RSVP stream. In the example shown here, observers have to report the numerical value of a target digit ('2') that is defined by a unique colour (red). When evidence accumulation about the target-defining colour reaches the engagement threshold, an attentional episode is triggered. This affects the processing of the target's identity, but also the activation profile of the post-target distractor ('7') that is processed within this episode, and thus also benefits from amplification. As a result, the post-target distractor may be encoded in WM alongside the target, or even instead of the target.

The control of attentional episodes by spatial, temporal, and feature-based attention

According to the UDAS, the central selective function of attention (controlling access of sensory representations to working memory and response selection) is implemented within attentional episodes. These episodes are triggered once parallel sensory evidence accumulation has reached a threshold value. This threshold serves as a tripwire that immediately and indiscriminately elicits the amplification of sensory representations at particular locations during the attentional episode. In other words, these episodes themselves are not under direct voluntary control, but are an automatic result of above-threshold activation levels. This may seem counterintuitive, since object selection is clearly goal-sensitive. Observers usually select objects and events that are relevant in a given task context. This task sensitivity of attentional episodes results from the fact that parallel

sensory evidence accumulation does not take place in an entirely bottom-up stimulus driven fashion, but is modulated by earlier attentional processes that operate prior to the start of the attentional episode, and are sensitive to top-down task goals. These attentional mechanisms (spatial attention, temporal attention, and feature-based attention) have been investigated intensively, but they have typically been understood as different forms of attentional selection. The UDAS proposes a different and more precise account of the role of these mechanisms by describing them as *precursors* of the selective amplification of processing that take place during attentional episodes.

Based on explicit instructions, prior experience or implicit learning (e.g., Chun & Jiang, 1999), participants can form expectations about task-relevant locations, time points or features. These specify where or when a relevant event object is likely to appear, and which features signal its arrival. For example, when waiting for a stop light to turn green, observers know where to attend, what colour to look for, and when the light is likely to change (based on the time that passed since it turned red). Advance information about the expected location (spatial attention), the defining features (feature-based attention), or the expected time of arrival (temporal attention) of relevant objects is represented in attentional templates (e.g., Duncan & Humphreys, 1992). Once activated, these templates modulate parallel sensory evidence accumulation processes by expediting and amplifying the processing of stimuli that match one or several of these attributes. Take for example a rapid serial visual presentation (RSVP) task where multiple items are presented sequentially and participants have to find a specific target among distractors. Evidence accumulation for items in the stream is facilitated by spatial attention, whereas only the target benefits from feature-based attention, and also benefit from temporal attention when its temporal position is predictable. This prioritization of template-matching stimuli increases the likelihood that they will cross the engagement threshold and thus trigger an attentional episode. As a result, these stimuli are more likely than non-matching stimuli to gain access to subsequent memory- and response-related stages.

While spatial, temporal and feature-based attention on the one hand and attentional episodes on the other all modulate sensory processing, they differ in several important aspects. Most importantly, their temporal characteristics are distinct. Spatial, temporal, and feature-based attention operate during early parallel stages, prior to the start of attentional episodes. They can also be sustained for extended periods, and can already be activated during the preparation for a task, prior to the arrival of external stimulation. In contrast, attentional episodes have transitory dynamics, are triggered typically around 170 ms after the arrival of a potentially task-relevant event (see below), and usually dissipate within another 100–200 ms. In addition, the sustained effects of
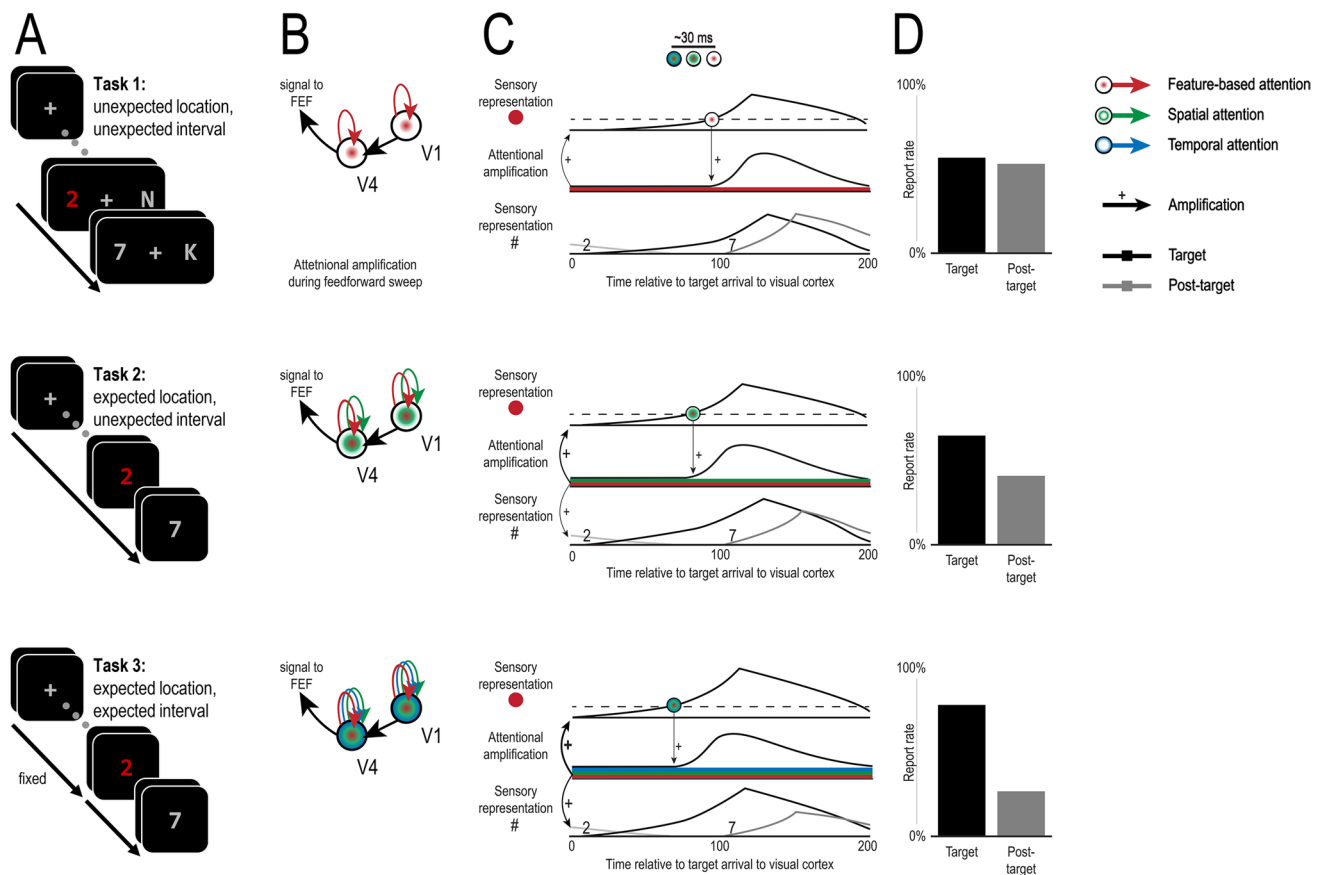
Fig. 2 Schematic outline of the time course of attentional modulations during early sensory processing on subsequent attentional episodes in three rapid serial visual presentation (RSVP) tasks that differ in spatial and temporal expectation regarding the target's appearance (**a**). Preparatory attentional templates representing target colour (feature-based attention), location (spatial attention), and onset time facilitate early stages of target processing during the feedforward sweep (**b**), resulting in faster evidence accumulation at the location of the target (**c**). The earlier the attentional episode, the higher the activation of the target relative to the post-target distractor, resulting in higher proportions of correct responses and fewer intrusion errors (**d**)

spatial, temporal, and feature-based attention on evidence accumulation are relatively subtle, whereas attentional episodes are associated with a rapid and large, albeit transient, increase in activation levels. Finally, while the former types of attention are sensitive to the task-relevance of locations, features, and time points, attentional episodes, once triggered, do not show this type of goal selectivity, and therefore can result in an indiscriminate amplification of both relevant and irrelevant objects.

Figure 2 illustrates this interplay between feature-based, spatial, and temporal attention, and subsequent attentional episodes, as described in our diachronic account of attentional selectivity. It shows the time course of attentional episodes in three RSVP tasks with color-defined targets that differ in spatial and temporal expectations: (i) a dual-stream RSVP task where the target's location and temporal position is unknown; (ii) a single-stream RSVP task where the target's location is known but the temporal position is not; (iii) a single-stream RSVP task where the temporal position

is predictable. In all three tasks, preparatory feature-based attention (tuned to the known target colour) facilitates target processing at an early point in time, even before its numerical value becomes available. However, processing of the target in these tasks differs in whether it also benefits from sustained spatial attention (task 1 vs. tasks 2 and 3) and from temporal attention (tasks 1 and 2 vs. task 3). While these early attentional modulations are not necessary for an attentional episode to be triggered (i.e., targets can also elicit an episode when their spatial and temporal positions are not predictable), they expedite the evidence accumulation process, and thus the point in time where the engagement threshold is crossed, and an attentional episode is triggered. Accordingly, the balance of activation between the target and post-target is more biased towards the target in task 2 relative to task 1 and in task 3 relative to task 2, leading to more correct responses and fewer erroneous reports of the post-target distractor (distractor intrusions). According to UDAS, a difference of a few dozen milliseconds in the onset of the

attentional episode should result in substantial differences to the ratio of correct responses versus distractor intrusions (e.g., Zivony & Eimer, 2021a).

Underlying neural mechanism: Recurrent processing

A poignant criticism of Broadbent's filter model (Allport, 1993) and indeed of many current cognitive models of selective attention is that they do not seriously engage with current neuroscientific findings (Allport, 1993; Di Lollo, 2018). Because our proposed account of visual selectivity emphasizes the temporal dynamics of attentional processes, it can be readily linked to the time course of their underlying brain mechanisms. In this section, we outline the possible neural basis of attentional episodes and their control by preceding attentional mechanisms. These suggestions are also illustrated in Fig. 1f.

Once a stimulus appears in the visual field, a retinotopically mapped signal from the retina reaches the visual cortex within approximately 50–80 ms, as indicated by successive early event-related potential components originating from primary (C1; Di Russo et al., 2002) and ventral extrastriate visual cortex (P1; Luck, Chelazzi, et al., 1997a). Shortly after, the signal reaches frontal areas, such as the frontal eye fields (FEFs) in prefrontal cortex (e.g., Ogawa & Komatsu, 2006). During this feedforward sweep, visual processing is affected by both 'bottom-up' factors, such as saliency, and 'top-down' factors, such as pre-established attentional templates and focused spatial attention (see Katsuki & Constantinidis, 2014, for review). Evidence for such early modulations comes from both single-cell and electrophysiological studies. Single-cell recordings demonstrate that when the target is readily discriminable, neural firing rate is enhanced for salient stimuli, stimuli at attended locations, and stimuli that match the attentional template within 120 ms after stimulus onset (e.g., Pooresmaeili et al., 2014). These findings suggest that during the feedforward sweep, neural activity is already affected by competitive processes (Desimone & Duncan, 1995) that affect the firing rate of neurons (Katsuki & Constantinidis, 2014), modulate neural gain (Itthipuripat et al., 2014), and tune neuronal activity to specific features (Martinez-Trujillo & Treue, 2004; Sani, Santandrea, Morrone, & Chelazzi., 2017). Electrophysiological studies also show clear evidence of attentional modulations during this early time window. Specifically, while the C1 has often been found to be unaffected by attentional manipulations (but see Slotnick, 2018), the subsequent P1 is enhanced when a visual stimulus appears in an attended location, an expected moment in time, or matches a target-defining feature (Luck et al., 1997a, b; Seibold et al., 2020, b; Warren, Yacoub, & Ghose, 2014; Zhang & Luck, 2009). These findings show that spatial, temporal, and feature-based attention can facilitate the neural processing during this early phase (Fig. 1f, left panel), presumably due to a sustained baseline shift in neural activity that is triggered prior to stimulus onset, during attentional preparation.

Feedforward processing yields sensory representations of limited strength and duration (Seth & Baars, 2005). At this stage, temporally adjacent objects can suppress the neural activity related to a particular object (Keysers & Perrett, 2002), thereby disrupting evidence accumulation and preventing its encoding in WM. Therefore, subsequent stages of attentional selectivity are needed. These stages involve feedback loops that result in recurrent processing between the visual cortex and higher brain regions. Evidence for recurrent processing comes from single-cell studies in monkeys showing that a visual stimulus evokes two separate waves of activity in the visual cortex (e.g., Lamme & Roelfsema, 2000) and from studies showing that disruption to processing in visual areas via transmagnetic stimulation occurs at different time windows (e.g., Camprodon, Zohary, Brodbeck, & Pascual-Leone, 2010). Top-down recurrent signals are thought to facilitate the activation states of some of the sensory representations that are generated during the feedforward sweep, thereby increasing the likelihood that they will be encoded in WM (Sligte, Scholte, & Lamme, 2008) and consciously perceived (Boehler et al., 2008; Lamme & Roelfsema, 2000). These feedback loops further modulate processing in the visual cortex, including midtier cortical areas in the ventral visual pathway, such as V4 and adjacent areas. Previous research has implicated the subcortical pulvinar nucleus in the thalamus as the origin of recurrent signals that subsequently modulate activation levels in V4 (Zhou, Schafer, & Desimone, 2016). However, the most prominent contribution to recurrent processing in V4 comes from the FEF.

The FEF plays a crucial role in guiding covert shifts of attention, the execution of eye movements (see Vernet, Quentin, Chanes, Mitsumasu, & Valero-Cabré, 2014, for review), and in our model, in triggering attentional episodes. While these control processes overlap (see *Attentional episodes and eye movements* section below), they are not identical. Previous studies suggested that the FEFs serve as a retinotopically coded saliency map (Katsuki & Constantinidis, 2014), where sensory signals processed during the feedforward sweep coalesce, and are translated into goal-related signals (Ibos, Duhamel, & Hamed, 2013; Ogawa & Komatsu, 2006). According to our current understanding, once activation within an FEF map reaches a threshold level, recurrent signals are projected to V4 (among other areas), resulting in a sustained processing amplification and corresponding locations. Thus, early attentional modulations of neural activity in the visual cortex likely contribute to recurrent processing by promoting above-threshold activity in FEF (Zhou & Desimone, 2011).

According to UDAS, recurrent activity from the FEF is the neural basis of the amplification of stimulus processing

during attentional episodes. Support for this hypothesis comes from studies showing that attentional modulations in FEF precede later attentional modulations in V4 (Cohen, Heitz, Schall, & Woodman, 2009; Purcell, Schall, & Woodman, 2013; Zhou & Desimone, 2011). Studies that used neurodisruption techniques also reveal the relationship between FEF and attentional modulations in V4 (see Chambers & Mattingley, 2005, for review): of the FEF results in enhanced activity in V4 whereas TMS-induced interference with FEF activity can result in deficits in attentional shifting. As a direct neural marker of an attentional episode having been triggered, neural activity in V4 spikes after approximately 150 ms (Cohen & Maunsell, 2009; Sani et al., 2017).

Recurrent signals that induce amplification of visual activation are selectively elicited in response to stimuli with currently task-relevant attributes. However, we suggest that their effects on visual processing within a given retinotopic area are indiscriminate, resulting in the amplification of all sensory input that is processed during the same attentional period. This hypothesis is supported by observations that an attentional increase in neural gain is not necessarily tuned to specific features (Baluch & Itti, 2011; Noudoost, Chang, Steinmetz, & Moore, 2010; Tootell et al., 1998). Therefore, when another stimulus is presented in close temporal proximity to a target, its feedforward activation level is increased, even when it does not have any target-matching attributes. This indiscriminate activation increases the likelihood that both the target (which triggered the attentional episode) and the temporally proximal item will be encoded, but also that competitive interactions between them are increased (Keysers & Perrett, 2002; Marti & Dehaene, 2017; Tang et al., 2020; see *Perceptual competition and exclusion from WM* section below).

In our model, recurrent signals from FEF are triggered when the activation level of a particular visual representation reaches the engagement threshold (see Fig. 1f, middle panel). These signals are sent to specific locations within retinotopic ventral visual areas, resulting in an attentional episode during which the processing of all sensory representations at this location is amplified (Fig. 1f, right panel). Critically, the modulations of sensory processing that occur during the early feedforward sweep as a result of spatial, temporal, and feature-based attention (see above) bias the probability that a specific feedforward activity can elicit a recurrent signal in favour of spatially or temporally attended stimuli and stimuli with target-matching features.

An electrophysiological marker of this activity amplification triggered by recurrent feedback signals is the N2pc (N2-posterior-contralateral) event-related potential (ERP) component (Eimer, 1996; Woodman & Luck, 1999). The N2pc consists of a larger negative deflection in electrodes contralateral to the eliciting stimulus and emerges between 180 and 300 ms from stimulus onset at lateral posterior electrodes over ventral visual cortex. The N2pc has been

previously associated with activity in V4 (Hopf, Boelmans, Schoenfeld, Luck, & Heinze, 2004; Hopf et al., 2006; Luck et al., 1997a, b; Westerberg, Schall, Maier, Woodman, & Schall, 2021) that is triggered by a recurrent control signal generated within FEF (Cohen et al., 2009; Purcell, Schall, & Woodman, 2013). N2pc components are usually interpreted as a marker of the attentional selection of target stimuli in visual search displays. This characterization reflects the general assumption of standard accounts of attention that 'selection' is a discrete step in visual processing that takes place at a specific moment in time. In the context of the diachronic model proposed here, the N2pc is linked to a temporally extended process (amplified processing during the attentional window). The start of this process (attentional engagement), which is triggered by recurrent signals generated during earlier phases of attentional processing (evidence accumulation based on spatial, temporal, or feature-based attention), can be tracked in real time on the basis of N2pc onset latencies (Zivony & Eimer, 2021a; see also Kiss, Van Velzen, & Eimer, 2008; Zivony, Allon, Luria, & Lamy, 2018). This is illustrated by studies demonstrating that attentional manipulations that affect the speed of evidence accumulation also affect N2pc onsets. For example, the N2pc emerges earlier when a target appears in a location that is already attended (Foster, Bsales, & Awh, 2020; Zivony & Eimer, 2021a, Experiment 3), at the expected moment in time (Hackley, Schankin, Wohlschlaeger, & Wascher, 2007; see also Grubert & Eimer, 2018, 2020, for additional evidence that temporal expectations about search display onsets determine attentional engagement), or when target-defining features are easier to detect (Brisson, Robitaille, & Jolicoeur, 2007; Callahan-Flintoft & Wyble, 2017; Töllner, Zehetleitner, Gramann, & Müller, 2011; Zivony & Eimer, 2021a). In addition to N2pc onset latencies, the amplitude of N2pc components is also affected by attentional factors that affect performance, including the reward history of attended objects (e.g., Bachman et al., 2020; Kiss, Driver, & Eimer, 2009). Such N2pc amplitude modulations may reflect the amount of amplification of specific objects during the attentional episode. If N2pc onsets and N2pc amplitudes reflect the start and the degree of attentional amplification, respectively, it is important to track the effects of attentional manipulations on both of these variables independently (e.g., Bachman et al., 2020; Zivony & Eimer, 2021b).

Evidence for the existence of attentional episodes

According to the UDAS, selectivity is the result of amplified perceptual processing during attentional episodes. The critical question is of course whether these episodes exist, and whether they operate as described in our account. Although

the importance of attentional episodes for perceptual selectivity has been emphasized in some computational models (Olivers & Meeter, 2008; Põder 2013; Shih, 2008; Smith & Wolfgang, 2004; Wyble et al., 2009, 2011;), they have not featured prominently in mainstream thinking about visual attention. A likely reason for these is that many contemporary models of visual selectivity are based on visual search tasks that use static displays (Treisman, 2014; Wolfe, 2014). Performance in such tasks can be accounted for by the difficulty of discriminating target and nontarget features, without reference to an extended period of attentional amplification that is triggered once target-defining features have been detected. Thus, while standard models of attention can explain how competitive interactions can affect attention to multiple areas or items in space (e.g., Benoni & Tsal, 2012; Duncan & Humphreys, 1992; Lavie, 1995; Theeuwes, 2010; Treisman, 1998, 2014; Wolfe, 2014, 2021), they do not take into account how attention unfolds in time and how this affects stimuli presented at multiple time points. More direct evidence for the existence of attentional episodes can be obtained in experimental paradigms that include the sequential presentation of multiple displays, and are thus more sensitive to the temporal aspects of attentional selectivity. In this section, we discuss two such paradigms – RSVP tasks, where streams of objects appear in rapid succession in the same location (see Fig. 3a–c), and the contingent-capture cuing task, where target objects are preceded by an object (cue) with target-matching features, either at the same or at a different location (see Fig. 3d and e). We review results obtained with these paradigms that provide evidence for two of the three premises of the UDAS: the existence of a threshold for attentional engagement and the principle of indiscriminate amplification of processing during attentional episodes. The first premise of UDAS (perceptual processing is an evidence accumulation process) is relatively uncontroversial (e.g., Ratcliff, Smith, Brown, & McKoon, 2016), and will therefore not be discussed in this section.

In a typical RSVP task, participants have to identify an alphanumeric target that is embedded in a stream of multiple distractors, presented at rapid rates (usually about ten items per second). When the target's response dimension is different than that of the distractors (e.g., a target digit among letters, Fig. 3a), it can be easily differentiated from the distractors, in spite of the fact that its temporal position in the stream is unpredictable (Potter, Wyble, Hagmann, & McCourt, 2014). Variations of this paradigm produce two highly reliable patterns of results. First, when the distractors share the target's response dimension (e.g., when the target is a digit in a stream of other digits, Fig. 3b), distractor intrusions are frequent: Participants will often erroneously report not the target, but the identity of a distractor that appears in close temporal proximity to the target (Botella, Suero, & Barriopedro, 2001; Chun, 1997; Goodbourn & Holcombe,

2015; Vul et al., 2008; Zivony & Eimer, 2020, 2021a). Second, when participants are asked to identify two successive targets (T1 and T2) in the RSVP stream (Fig. 3c), the accuracy of reporting both targets is high when they appear more than half a second apart, but report accuracy for T2 is strongly impaired when it is presented within 200–500 ms after T1 (attentional blink; Raymond, Shapiro, & Arnell, 1992; see Zivony & Lamy, 2021, for a discussion of different accounts of this phenomenon).

There are many different types of attentional cueing procedures. In some of these tasks, cues are salient events (e.g., abrupt onsets) that are followed by targets at cued or uncued locations, and the temporal interval between cues and targets is manipulated. Performance benefits for cued targets are usually maximally around 100–200 ms after cue onset, and then either remain constant (Cheal & Lyon, 1991; Cheal, Lyon, & Hubbard, 1991), or gradually decrease (Nakayama & MacKeben, 1989; Müller & Findlay, 1988; Wilschut, Theeuwes & Olivers, 2011, 2013, 2015). Such findings suggest that these cues facilitate target processing within a temporally extended attentional episode, as suggested by our UDAS model. However, since the cues used in these studies typically did not share any target features, these episodes are presumably not the result of attentional engagement processes guided by target templates, but are instead triggered primarily by bottom-up salience signals (see *Evidence for an engagement threshold* section for further discussion). For this reason, we focus here primarily on the contingent-capture cueing paradigm (Folk et al., 1992; Remington & Folk, 1998) that has been employed to investigate involuntary attentional capture by task-irrelevant objects that share target-defining features. In such cuing tasks, search displays are preceded by cue displays, which participants have to ignore. Search targets are presented at unpredictable locations, and usually require a speeded discrimination response (e.g., whether the target is slanted to the right or the left, or whether it is an E or an H; see Fig. 3d and e). Cue displays presented prior to the search display contain an object (cue) that either matches or does not match the target's defining feature (e.g., its colour), but is spatially uninformative about target location (i.e., the target is equally likely to appear at cued and uncued locations). Location benefits (better performance for targets and cued as compared to uncued locations) observed with target-matching cues in this paradigm are interpreted as evidence that such cues involuntarily capture spatial attention (e.g., Eimer & Kiss, 2008; Folk & Remington, 1998; Goller, Schoeberl, & Ansorge, 2020; Zivony & Lamy, 2018). Because these cues match the currently active attentional search template, they activate top-down control processes that guide attention to their location, even though this does not provide any strategic benefits for subsequent search performance. Non-matching cues typically do not elicit any location benefits, even when they are physically
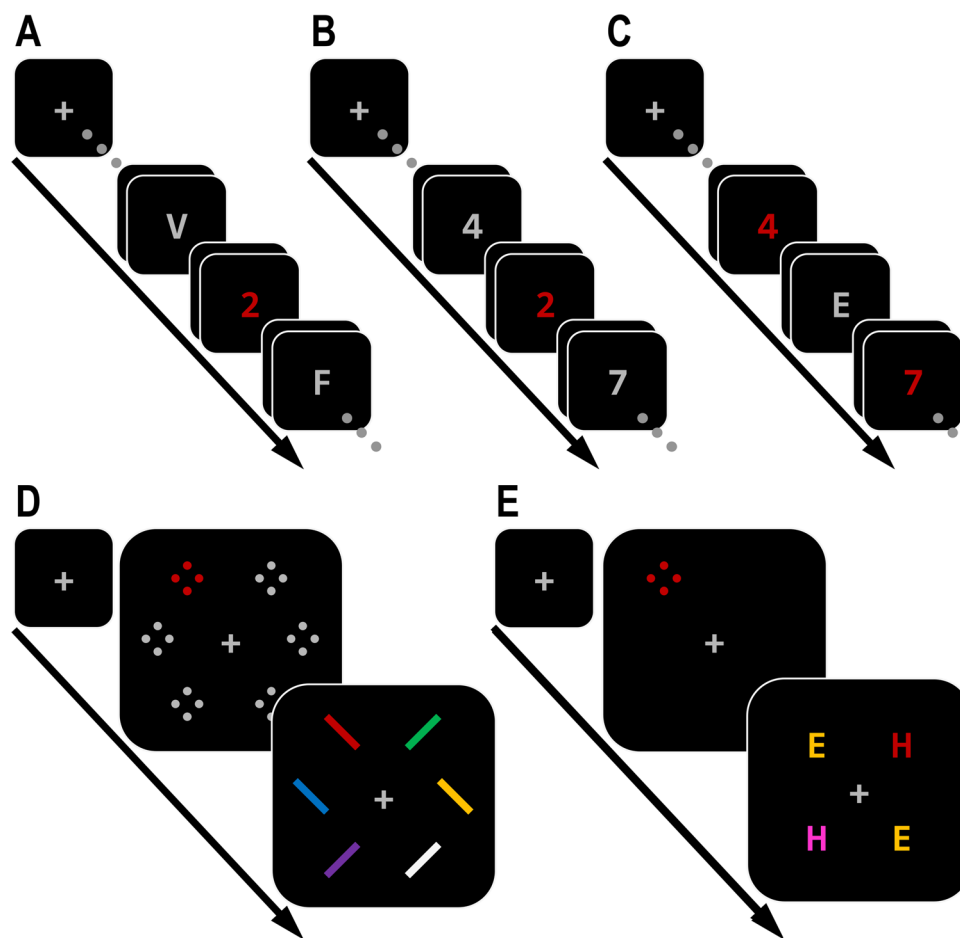
Fig. 3 Illustrations of rapid serial visual presentation (RSVP) and contingent capture cuing experiments. RSVP experiments (top row) include multiple frames separated by blank screens, presented sequentially at a typical rate of approximately ten items per second. In the examples presented here, the task is to report without time pressure the (**a**) red digit among grey letters, (**b**) red digit among grey digits, and the (**c**) two red digits among grey letters. In contingent capture cuing experiments (bottom row), participants usually make a speeded response to a target defined by a specific feature (such as colour), preceded by a single cue display that includes an irrelevant singleton distractor that either matches or does not match the target-defining feature. In the examples presented here, the task is to report as fast as possible the (**d**) orientation of the red bar, and (**e**) whether the red letter is E or H. In (**d**), the target-matching cue appears in the location of the target. In (**e**), this cue appears in the location of a distractor that is response-incompatible with the target

salient, demonstrating that this type of attentional capture is contingent on task settings.

In the next two sections, we review results from RSVP and contingent-capture spatial cuing tasks that support the second and third premise of the UDAS. We start with presenting evidence for indiscriminate amplification during attentional episodes, as this most clearly distinguishes between standard temporally discrete selection concepts and diachronic accounts of attentional selectivity. Then we will review evidence in support of an engagement threshold.

Evidence for indiscriminate amplification

According to standard accounts of attentional selectivity, the efficiency of selecting a particular target among distractors is determined by how easily these two types of items can be distinguished. In visual search and RSVP tasks where the target is defined by a distinct perceptual feature (e.g., a specific colour), target selection should therefore be straightforward, and target reports highly accurate (e.g., Chun & Potter, 1995; Di Lollo et al., 2005; Taatgen, Juvina, Schipper, Borst, & Martens, 2009; Wolfe, 2021). In other words, there is no reason to assume that distractors are frequently encoded and reported under such conditions. In contrast, the UDAS framework proposes that the detection of a target-matching feature does not immediately result in the selection of the object to which this feature belongs, but instead triggers a temporally extended attentional window, during which targets and distractors are preferentially processed in an indiscriminate fashion. Evidence for this indiscriminate amplification cannot easily be obtained in single-frame visual search experiments, but is abundant in attentional

paradigms such as spatial cueing or RSVP where two or multiple stimulus frames are presented in rapid succession. Indiscriminate amplification substantially increases the likelihood that several successive items are encoded. Under certain circumstances, this can result in distractor intrusion errors, where one of the distractors will be reported instead of the target.

Clear evidence for the encoding of multiple items that appear in close temporal succession in RSVP streams comes from the phenomenon of lag-1 sparing in RSVP streams with two targets. In these tasks, the attentional blink is either partially or entirely absent when T2 immediately follows T1 (e.g., Potter, Chun, Banks, & Muckenhoupt, 1998). The same is true when RSVP streams include three targets, where the third target (T3) can be spared from the blink if it is preceded by T2 (Dell'Acqua, Dux, Wyble, & Jolicoeur, 2012; Kawahara, Kumada, & Di Lollo, 2006; Olivers, Van der Stigchel, & Hulleman, 2007). T3 sparing does not depend on correct T2 reports. For example, Kawahara et al. (2006) showed that even when T2 appeared inside the blink period (and was missed itself), T3 accuracy was improved on when T3 followed T2 immediately than with longer T2-T3 lags. In addition, report accuracy for a target item within the blink period is also improved considerably when it appeared immediately after a target-matching cue (Nieuwenstein, 2006; Nieuwenstein, Chun, van der Lubbe, & Hooge, 2005; Zivony & Lamy, 2016b).

While such observations provide evidence for the encoding of multiple successive items within RSVP streams, they do not in themselves demonstrate indiscriminate amplification during attentional episodes, since the items encoded in these attentional blink experiments were all targets. It is thus plausible to assume that lag-1 sparing occurs because the input filter that selects target-matching sensory input remains tuned to the target's category (e.g., Di Lollo et al., 2005; Taatgen et al., 2009). This explanation maintains the temporally discrete concept of selection by postulating that multiple items can trigger separate selection processes in close temporal succession when they all match the current attentional template. In contrast, if amplification was genuinely indiscriminate, it should not only facilitate target encoding and reports, but should also equally facilitate the processing of task-irrelevant distractors that appear in close temporal proximity to targets (or of objects with target-defining features). This has indeed been shown in numerous experiments with RSVP streams that only included a single feature-defined target. As mentioned previously, participants often commit distractor intrusion errors in such tasks, and these errors are sometimes more frequent than correct target reports. Critically, and conceptually parallel to lag-1 sparing, such distractor intrusions show a distinct temporal pattern: Participants are much more likely to report the distractor that immediately follows the target, and only rarely report

distractors that precede the target (e.g., Botella et al., 2001; Goodbourn & Holcombe, 2015). For example, consider a study where the target is a red digit embedded among grey digits, and the target is a red '2', preceded by a grey '4' and followed by a grey '7' (Fig. 3b). In this case, participants are much more likely to report seeing a red '7' than a red '4'. In other words, in both attentional blink and distractor intrusions studies, the presence of a target increases the likelihood that the immediately following item will be encoded and reported.

Analogous findings come from two-target attentional blink studies that also measured distractor intrusions following T2. Here, post-T2 intrusions errors also occur, and these errors are even more frequent when T2 appears during the blink than when it appears outside the blink period (e.g., Chun, 1997; Goodbourn & Holcombe, 2015; Vul et al., 2008). Furthermore, while cuing by a target-matching object can increase target reports, it can also in an increase of distractor intrusion errors. For example, in an attentional blink experiment by Zivony and Lamy (2016b), where T1 and T2 were colour-defined letters, T2 reports during the blink period were more accurate when T2 was immediately preceded by a colour-matching cue, but not with nontarget-colour cues. Critically, the same effect was also observed for distractor intrusions errors, which were much more frequent when distractors were preceded by colour-matching cues. These findings cannot be explained by the hypothesis that multiple temporally discrete template-guided selection processes are elicited successively (Di Lollo et al., 2005; Taatgen et al., 2009), because distractors did not match the attentional template, and should thus not be selected and encoded in WM. Overall, these observations provide strong evidence that the detection of a target or a target-defining feature in RSVP tasks triggers an attentional episode during which the processing of targets as well as distractors is indiscriminately enhanced.

Along similar lines, location effects in contingent-capture spatial cuing tasks suggest that items at cued locations benefit from amplified processing during the attentional episode that is triggered by a target-matching cue stimulus. However, as such effects are typically measured for performance in response to search targets, they do not provide direct evidence of indiscriminate amplification. Such evidence comes from the presence analogous location effects for distractors at cued locations. Many studies have demonstrated that such distractors elicit reliable response compatibility effects on target performance (e.g., faster RTs when the cued distractor and the target are mapped to the same as compared to different responses; Carmel & Lamy, 2014; Remington & Folk, 2001; Maxwell, Gaspelin, & Ruthruff, 2020; Zivony & Lamy, 2016a, 2018; see also Chen, Leber, & Golomb, 2019, for a similar paradigm). In contrast, no such compatibility effects were found for distractors that

follow target-nonmatching cues. The presence of compatibility effects demonstrates that distractors were processed up to the level where their identity was encoded (Eriksen & Eriksen, 1974). According to the UDAS account proposed here, this is the result of the fact that these distractors were processed within the attentional episode triggered by preceding target-matching cues.

Standard accounts of attentional selectivity may account for distractor intrusions and distractor compatibility effects by assuming that the detection of target-defining features and the selection of objects for WM encoding are entirely separate processes. For example, McLean et al. (1983) suggested that when target detection is strongly delayed in single-target RSVP tasks, the item that follows the target is selected for report. From this temporally discrete account, it follows that the number of items selected for further processing should not exceed the number of targets items that have to be detected. In contrast, the indiscriminate amplification hypothesis predicts that multiple successive items can be amplified and encoded when they are presented during a single attentional episode, and that this is the case regardless of whether only a single target object or multiple targets have to be found. Thus, in visual search and RSVP tasks where observers search for a single target, there is no reason for participants to select and encode more than one object. Evidence for the encoding of multiple items in such tasks would therefore provide further evidence against the standard discrete account of selection.

In two-target RSVP attentional blink tasks, the lag-1 sparing effect is assumed to occur because T1 and T2 are both encoded within the same attentional episode. Electrophysiological evidence for this comes from observations that often only a single N2pc component (a marker of attentional engagement, see above) is elicited when T1 and T2 appear in immediate succession (Tan & Wyble, 2015: see also Akyürek & Meijerink, 2012; Callahan-Flintoft, Chen, & Wyble, 2018; Zivony & Eimer, 2020). This suggests that the attentional engagement triggered by T1 is sufficient for both targets to gain access to WM. However, and critically, if amplification during attentional episodes was genuinely indiscriminate, multiple successive items should also be encoded together in single-target RSVP tasks. The ubiquity of distractor intrusion errors in such tasks does not in itself demonstrate multiple-item encoding, since it is possible that these errors result from encoding the post-target distractor instead of the target in working memory. However, several RSVP studies have shown that more than one item is often encoded in such tasks. For example, Vul et al. (2009) employed RSVP letter streams, where the target was defined by an enclosing circle, and participants had to provide two guesses about the target's identity. Results showed that participants very often reported more than one item from the RSVP stream, without resorting to guesses. More recently,

we (Zivony & Eimer, 2021a, b) used similar procedures, and employed both behavioural and electrophysiological markers of WM storage (CDA components; Vogel & Machizawa, 2004; see also Luria, Balaban, Awh, & Vogel, 2016, for review), to demonstrate that both the target and the post-target distractor are often encoded in WM, even in tasks where participants are asked to provide only a single response.

Analogous conclusions have also been drawn from results of studies using the contingent capture cuing task. These studies showed that target-matching cues do not merely facilitate the selection and encoding of subsequent items at cued location, but are also often encoded (Chen & Wyble, 2018; Maxwell et al., 2020). For example, Maxwell et al. (2020) used cues that were themselves associated with a target-compatible or incompatible response. Target-matching cues as well as cues that were predictive of the target's location produced both a location benefit and a response compatibility effect for subsequent search targets. These results indicate that both the cue and the target were encoded and identified during the same attentional episode, even though participants always searched for a single target object.

Evidence for an engagement threshold

According to Premise 2 of UDAS, an attentional episode is triggered only after sufficient evidence about the presence of a potentially task-relevant stimulus is accumulated. This premise includes two claims. First, attentional engagement is only triggered by visual signals that match currently active top-down goals. Second, the speed with which engagement processes are elicited depends on perceptual processing of these signals prior to the start of any given attentional episode. The onset of attentional episodes should therefore be different for different types of signals. Even for a single constant target-defining signal, trial-by-trial fluctuations in signal extraction processes should result in variations in the speed with which such episodes are triggered. In this section, evidence for both these claims is discussed.

If engagement is contingent on top-down goals, stimuli that do not match the current search template should not trigger an attentional episode. This argument is supported by the work of Folk and colleagues (Folk et al., 1992; Folk & Remington, 1998), who showed that attentional capture (which precedes attentional engagement) is mediated by task goals and is therefore triggered only by stimuli with target features, but not by stimuli without such features, even when they are perceptually salient. Evidence for this claim come from spatial cuing studies, where location benefits are only observed for cues that share target features (e.g., Folk & Remington, 1998; Lamy, Leber, & Egeth, 2004), and from ERP studies, where only target-matching cues trigger N2pc components indicative of attentional capture (e.g., Eimer &

Kiss, 2008; Goller et al., 2020). Abrupt onsets may however be a possible exception to this general rule (e.g., Darnell & Lamy, 2021; Gaspelin, Ruthruff & Lien, 2016; Lamy, Darnell, Levi, & Bublil, 2018; Maxwell et al., 2020; Zivony & Lamy, 2018). For example, Gaspelin et al. (2016) showed that in difficult visual search tasks, task-irrelevant abrupt-onset cues produced reliable location benefits indicative of attentional capture (but see Darnell & Lamy, 2021, for a different interpretation). However, attentional capture by irrelevant abrupt onsets does not necessarily imply attentional engagement. Zivony and Lamy (2018) demonstrated that both task-relevant and task-irrelevant abrupt-onset cues produced location benefits (indicative of attentional capture), but only the task-relevant cues produced distractor compatibility effects (indicative of attentional engagement). Analogous results were found by Maxwell et al. (2020), who also found distractor compatibility effects for cues that did not match any target features, but only when they were predictive of the target's location. These findings suggest that attentional engagement is strictly dependent on top-down goals, whereas attentional capture may also be triggered by bottom-up saliency signals alone under some conditions. In line with this conclusion, Goller et al. (2020) demonstrate that abrupt onsets trigger N2pc components only in contexts where they are task-relevant. Overall, these results support the hypothesis that attentional capture and subsequent attentional engagements are mediated by and contingent on current top-down task goals.

According to premise 2 of UDAS, attentional episodes are not triggered immediately once the potentially relevant information reaches visual processing areas, but only after the relevant sensory evidence crosses a hypothetical activation threshold. If this was the case, task-relevant sensory attributes that are processed faster should cross the engagement threshold earlier and therefore trigger attentional episodes more rapidly. Furthermore, since perceptual processing and evidence accumulation are inherently noisy processes that fluctuate across trials, the point in time when this threshold is reached should be variable. Importantly, this variability should also have measurable effects on performance in attentional selection tasks. Evidence for both predictions come from ERP studies that focus on the onset latency of N2pc components as markers of the start of attentional engagement. First, numerous studies have shown that N2pc onset latencies to targets in visual search and RSVP tasks depend on the speed with which target-defining features can be detected. Features that are processed faster, such as colour singletons, produce earlier N2pc than features that take longer to be processed, such as non-singletons and shapes (Brisson et al., 2007; Callahan-Flintoft & Wyble, 2017; Töllner et al., 2011; Zivony & Eimer, 2021a, Experiment 2). Similarly, directing attention to the location of an upcoming target with spatial cues or

verbal instructions should expedite evidence accumulation relative to an unfocused attention condition. Target N2pc components do indeed emerge earlier when the target's location is known in advance (Foster et al., 2020, Zivony & Eimer, 2021a, Experiment 3), indicating that attentional engagement is triggered more rapidly. Evidence for trial-by-trial variability in the speed of evidence accumulation and the onset of attentional engagement comes from studies that measured N2pc onset latencies in different types of trials that were separated by behavioural markers. In visual search studies, target N2pcs not only emerge earlier for participants who show fast responses to targets than for those with slower responses, but also within participants for trials with fast versus slow response times (Drisdelle, West, & Jolicoeur, 2016; see also Eimer & Mazza, 2005, for similar N2pc onset differences in a change detection task). In RSVP tasks, trials with post-target distractor intrusion errors are associated with delayed N2pc onsets relative to trials with correct responses (Zivony & Eimer, 2020, 2021a, b). This N2pc onset difference remains even when attention is focused in advance on the target RSVP stream (Zivony & Eimer, 2021a, Experiment 3), demonstrating that it does not reflect variability in the speed of attentional shifts. These observations are important, because they demonstrate that trial-by-trial variability in attentional engagement can have a strong impact on which items in RSVP streams are amplified, identified and encoded, and thus on the content of perceptual reports.

Overall, the studies reviewed in this section suggest that the speed of attentional engagement varies as a function of the speed with which different task-relevant features are processed, and also on a trial-by-trial basis, due to noise in the evidence accumulation process. While these findings support the notion of an engagement threshold, our UDAS account does not at present specify the level of accumulated evidence that is sufficient for an attentional episode to be triggered, and whether the engagement threshold is fixed or can be adjusted in line with task demands or current contexts.

One interesting possibility that needs to be explored in future work is that selection history plays a critical role in modulating this threshold. Numerous recent studies have demonstrated that the history of previous attentional deployments to particular targets has a strong impact on the efficiency of attentional selectivity. Targets that have been selected on recent trials are generally detected and identified better than new targets (e.g., Maljkovic & Nakayama, 1994; see also Awh, Belopolsky, & Theeuwes, 2012, and Failing & Theeuwes, 2018, for further discussion). This type of intertrial priming might be mediated by differences in attentional engagement thresholds, with lower thresholds for feature and objects that have triggered an attentional episode on preceding trials. Electrophysiological evidence for this

hypothesis comes from studies (e.g., Christie, Livingstone, & McDonald, 2015; Eimer, Kiss, & Cheung, 2010) that measured target N2pc components in a priming-of-popout task (Maljkovic & Nakayama, 1994) where targets were feature singletons (e.g., a red target among green distractors or vice versa). N2pc components emerged about 30–50 ms earlier on repetition trials where the target was the same as on the preceding trial than on change trials where a different target was presented. This observation demonstrates that intertrial priming modulates the speed with which attentional episodes are triggered. One plausible mechanism that can account for this effect is that selection history changes the prioritization of target features (Christie et al., 2015). In turn, this should modulate the speed of evidence accumulation at early parallel stages of visual processing, resulting in higher activation levels for sensory representations of repeated targets, which will therefore reach the (fixed) engagement threshold more rapidly. Alternatively, selection history may directly affect engagement thresholds, and lower the activation level required for engagement by repeated relative to new targets. While some evidence supports this latter proposition, research on priming-of-popout has not yet yielded a clear conclusion on whether intertrial priming affects guidance and early perceptual priming, independently from task goals (see Ramgir & Lamy, 2021, for review). Therefore, these two accounts of the link between selection history and the timing of attentional episodes will need to be dissociated in future research.

The costs of processing multiple items within the same attentional episode

The evidence reviewed so far suggests that multiple items can benefit from amplification during a single attentional episode, and are encoded together, regardless of whether they match the current attentional template. While this can improve performance when two task-relevant objects are presented in close succession (lag-1 sparing), it can also result in specific types of errors. In addition to the post-target distractor intrusions and distractor compatibility effects discussed before, there are other kinds of errors that arise when two items are processed within the same episode. These items may be perceived to appear in the wrong order (order reversals), fused together (temporal integration), and one of them may even be entirely excluded from encoding and conscious perception. Understanding how the processing of stimuli within attentional episodes can sometimes disrupt performance (instead of enhancing it) is crucial for informed interpretation of any future research guided by the diachronic view. Therefore, this section provides a brief review of these types of errors. Within the UDAS framework, they can be accounted for as results of temporal

variability of attentional engagement and competitive interactions between items within the same attentional episode.

Order reversals and temporal integration

Lag-1 sparing is generally regarded as beneficial in reducing or eliminating the attentional blink for T2, but this often comes at the price of impaired temporal order information. Even when participants correctly report both targets at lag-1, they are frequently reported in the wrong order (as T2 appearing before T1; e.g., Hommel & Akyürek, 2005; Potter, Staub, & O'Connor, 2002; Reeves & Sperling, 1986). Olivers, Hilkenmeier, and Scharlau (2011) suggested that attentional engagement is an important factor in determining which item enters into conscious perception first, in line with the prior entry literature (see Spence & Parise, 2010, for review). When an attentional episode is triggered by T1, the next item in an RSVP stream (T2) may receive more attentional enhancement. As a result, T2 may be processed and encoded faster on some trials, despite its later appearance. Evidence for links between the relative activation levels of T1 and T2 and order reversals was found by Olivers et al. (2011; see also Hilkenmeier et al., 2012a; Hilkenmeier et al., 2012b) in experiments where this activation was manipulated by cues that preceded either T1 or T2. Order reversals were reduced when T1 was cued and increased when T2 was cued. Within the UDAS framework, such order reversals are the result of the same factors that are responsible for distractor intrusion errors (Zivony & Eimer, 2021a, b): due to the inherent variability in the speed of attentional engagement (see *Evidence for an engagement threshold*), both types of errors occur on trials where engagement is slow, and thus facilitates the processing of the post-target item more than that of the target. With fast engagement, distractor intrusions are rare, and two successive targets are likely to be reported in the correct order.

A related phenomenon observed when two targets are presented in rapid succession is that they can become perceptually fused. Such fusion effects were demonstrated in a series of studies by Akyürek and colleagues (2012; Akyürek & Wolff, 2016; Akyürek, Akyürek, Kappelmann, Volkert, & Van Rijn, 2017; Simione, Akyürek, Vastola, Raffone, & Bowman, 2017). For example, when T1 and T2 are a forward leaning dash ('/') and a backward leaning dash ('\'), participants will often report seeing a single target - an 'X'. According to the temporal integration account (Hommel & Akyürek, 2005; Akyürek et al., 2012), the loss of temporal information can result in a fused percept of two (or more) successive items. This integration occurs when two items within a single attentional episode are similarly activated, which results in them being encoded in WM as a single perceptual event. When these items cannot be easily fused (as in the case of successive digits or letters), the absence of temporal order

information forces observers to guess which item was presented first. Again, according to UDAS, the presence of a single fused percept depends on the speed of attentional engagement. On trials with very fast or slow engagement, either the first or the second item will receive maximal enhancement, and this imbalance in activation makes it unlikely that fusion occurs. With intermediate engagement speeds, the relative activation of both items will be similar, which can result in them becoming perceptually fused.

Perceptual competition and exclusion from WM

While T2 accuracy is relatively high when T2 immediately follows T1 (lag-1 sparing), accuracy for both targets is usually somewhat lower in such situations than when both are presented 500 ms apart (i.e., outside the blink period). T1 accuracy is lower still when it is immediately followed by an especially salient T2 (e.g., Chun, 1997; Potter et al., 2002). Wyble et al. (2009, 2011) suggested that such results occur due to perceptual competition, which is biased in favour of more salient items, and can delay or even completely prevent the encoding of less strongly activated items in WM. This notion of perceptual competition is adopted in UDAS and is explained as the result of multiple objects being processed within the same retinotopic visual brain areas. In line with this proposal, Hilkenmeier et al. (2012a) showed that cuing T1 did not merely reduce order reversals, but also resulted in lower T2 accuracy (see also: Dell'Acqua et al., 2012). Analogous competitive effects have been reported in experiments where targets were immediately preceded by cue stimuli. Wilschut et al. (2013, Experiment 4) found that salient cues produced smaller benefits for target report accuracy than non-salient cues. Similarly, Pratt, Hillis, and Gold (2001) showed that cues that surround but do not physically overlap with the target produce larger performance benefits to reaction times than overlapping cues. Chen and Wyble (2018) demonstrated that cues are sometimes encoded in WM and that the likelihood of encoding the cue was negatively correlated with performance in identifying a subsequent target.

Overall, these behavioural findings suggest that competitive interactions between items within the same attentional episode can affect the ability of any of these items to be encoded within WM and become accessible to perceptual reports. In a recent study (Zivony & Eimer, 2020; see also Zivony & Eimer, 2021b), we provided more direct behavioural and electrophysiological evidence for this claim. In a task with two lateral RSVP streams, participants were allowed two guesses about the target's identity. When the target was followed by a potentially intruding distractor, participants sometimes encoded both the target and post-target distractor but were still less likely to report the target than on trials where it was not followed by such a distractor. Moreover, CDA components had a

larger amplitude when participants reported seeing both the target and post-target distractor relative to when they reported only one of these items. As the amplitude of the CDA is sensitive to the amount of information stored in WM (Luria et al., 2016; Vogel & Machizawa, 2004), this result indicates that the target was entirely excluded from access to WM on a proportion of trials, presumably due to early perceptual competition between the target and the post-target.

Perceptual competition may also be responsible for the phenomenon of object substitution masking (OSM). In a typical OSM experiment, a target is surrounded by four dots, and identification accuracy is impaired when the offset of the dots is delayed relative to target offset (Di Lollo, Enns, & Rensink, 2000). While OSM has been explained with reference to re-entrant processing (Di Lollo et al., 2000; Di Lollo, 2014), competitive interactions between two items within the same attentional episode may provide an alternative account. For example, Põder (2013) suggested that OSM can be interpreted as an attentional gating effect (Reeves & Sperling, 1986), where the surrounding dots benefit more from attentional enhancement than the target, due to their delayed offset. The UDSA framework offers a similar but more specific account. It assumes that the target and the dots are both processed within a single attentional episode triggered by the target, which amplifies the sensory processing of the dots. The dots' processing is also facilitated by its delayed offset relative to the target. This can bias perceptual competition in favour of the dots, and can prevent the target from being encoded in WM. As attentional episodes are assumed to be under the control of recurrent signals (see above), this account still retains the central proposal by Di Lollo (2014) that OSM is a phenomenon that is generated at a relatively late stage of re-entrant visual processing.

This explanation of OSM as a result of a competition for WM encoding within a single attentional episode is in line with the observation that the N2pc (a marker of attentional engagement triggered by the target) is unaffected by OSM (Woodman & Luck, 2003). Furthermore, the observation that the CDA component (a marker of WM storage, see above) in the OSM paradigm is larger on trials with correct target reports than trials where the target was missed (Salahub & Emrich, 2018) is in line with the suggestion that targets are reported when they are encoded together with the dots in WM, but are missed when they lose this competition for WM encoding.

The functions of attentional episodes

One reason why the standard account of attentional selectivity initially proposed by Broadbent (1958) was exceptionally influential was its ability to provide a succinct description of

the function of attentional selection. The selective filter was required at the interface of capacity-limited and capacity-unlimited processing stages in order to prevent informational overload. This assumption is reasonable in a linear serial stage model, but less so when the parallel and recurrent architecture of neural systems is taken into account (Allport, 1993). Still, the general issue about the functional role of attentional selectivity remains. For our diachronic account, which assumes that temporally extended attentional episodes are the core mechanism of selectivity, this question specifically relates to the function of these episodes.

We propose that the amplification of specific objects during attentional episodes probabilistically increases the likelihood that these objects will be encoded in working memory, in order to become accessible to other cognitive processes, and to flexibly guide action. Because attentional episodes are not triggered at random, but are mediated by earlier spatial, temporal, and feature-based attentional mechanisms, the objects that are selectively enhanced during these episodes are likely to be task-relevant. Critically, this selectivity is a consequence of differential activation of potentially relevant and currently irrelevant object representations that already emerges during the initial feedforward processing of visual input, and this differentiation is further amplified through recurrent feedback signals that trigger attentional episodes. In contrast to standard models, this account rejects any clear dichotomy between pre-attentive and attentive processing stages, and also the idea that capacity limitations are the root cause for attentional selectivity in perceptual processing. Instead, as discussed above, the amount of information that can be encoded and maintained in WM is mainly determined by the interaction between attentional enhancement and perceptual competition (see Oberauer & Lin, 2017, for a similar account).

The question remains why attentional episodes should have a specific duration, typically around 200 ms. Given the costs of temporally extended amplified processing reviewed previously (e.g., distractor intrusions, order reversals), would it not be more adaptive if these episodes were considerably shorter? First, it is important to note that these costs are usually observed in highly unusual lab-generated contexts, where individual objects are presented very briefly and follow each other in rapid succession. In more natural settings, even fast-moving objects are not usually replaced by entirely different objects within 100 ms. It is possible that there was no evolutionary pressure to develop a system that is able to generate and retain fully veridical representations of individual objects under such conditions. Thus, the costs that may occur in RSVP tasks may be more than offset by the multiple benefits of temporally extended attentional episodes. First, different object features are processed in different areas of the visual cortex at different speeds, potentially resulting in temporal binding problems mentioned earlier. It is likely

that attentional enhancement must persist for a sufficiently long period to allow for correct binding. In addition, WM encoding is itself a temporally extended process, which may require perceptual representations to retain their amplified activation states for long enough in order to be transformed into more stable WM representations (e.g., Brockmole, Wang, & Irwin, 2002).

On the other hand, there are also good functional reasons why the duration of attentional episodes should be limited. First, if attentional amplification during these episodes results in the recruitment of a large interconnected network of brain areas, this may be associated with high bioenergetic costs (Lennie, 2003). Therefore, it may be economical for each episode to be relatively short and only be triggered by potentially task-relevant input. Second, longer attentional episodes would inevitably increase competition between stimuli that are processed within the same episode. And finally, a sustained amplification of neural activity in retinotopic visual areas during longer attentional episodes may have adverse effects on both the sensory analysis of new visual input and/or the efficiency of WM maintenance, in particular if sensory encoding and WM storage utilize overlapping regions in visual cortex (Ester, Anderson, Serences, & Awh, 2013; Rademaker et al., 2019). Thus, the length of attentional episodes may reflect a functional compromise. They are long enough to ensure correct feature binding and successful WM encoding for potentially task-relevant objects, but short enough to limit adverse effects of competitive interactions between multiple objects and between sensory and WM representations.

In addition to facilitating the encoding of task-relevant objects in WM, attentional episodes may also have an important role for adaptive oculomotor control. It may be more than a coincidence that the typical duration of attentional episodes is similar to the typical duration of individual eye fixations during visual search and related tasks (e.g., Salthouse & Ellis, 1980). It is possible that one critical function of attentional episodes is to facilitate the extraction of task-relevant signals during the period when the eyes are stationary and retinal input is stable. Furthermore, attentional episodes could in principle also be involved in the selection of new saccade target objects. Several studies have shown that eye movements towards a target are preceded by N2pc components (e.g., Huber-Huber, Ditye, Fernández, & Ansorge, 2016; Weaver, van Zoest, & Hickey, 2017). This is not surprising since eye movements and attentional episodes share a neural substrate, as both rely on activation in the FEF (e.g., Moore & Fallah, 2004; Purcell et al., 2013; Vernet et al., 2014; Zhou & Desimone, 2011). However, associations between eye movements and N2pc components have not always been observed. For example, Talcott and Gaspelin (2021) recently demonstrated that eye movements towards a target were preceded by an N2pc only under

conditions where observers had to first covertly attend to a target stimulus in order to identify it as a Go or No-go stimulus for subsequent eye movements. In contrast, no pre-saccade N2pc was observed in a more conventional search task where observers could freely move their eyes within the search display and eye movements towards non-target objects were not explicitly discouraged. While the relatively short latency of eye movements in this task (about 200 ms after search display onset) may have made it difficult to detect clear N2pc components prior to saccade onset, these findings suggest that not all eye movements during search are preceded by an N2pc. Further evidence for this comes from the observation that search targets sometimes trigger an N2pc even when the initial saccade is directed towards a different nontarget object (Weaver et al., 2017). Such dissociations between saccade initiation and attentional episodes (as reflected by N2pc components) may be explained by assuming that both are based on overlapping evidence accumulation mechanisms, but that the threshold for eliciting attentional episodes is lower than the threshold for triggering eye movements. In line with this hypothesis, there is evidence that FEF stimulation can modulate attentional responses in V4 (indicative of recurrent processing) without triggering an eye movement (e.g., Moore & Fallah, 2004).

Another possible role of attentional episodes in the context of oculomotor control may be related to the rapid remapping of object features and locations that takes place when a saccade is planned and executed (Duhamel, Colby, & Goldberg, 1992). Remapping is reflected by the activation of neurons by objects in the anticipated (i.e., post-saccadic) receptive field even before the eye movement is initiated, and plays a role in the perceived constancy of object locations across saccades. However, such fast preparatory shifts of attention towards future fixation locations during saccade planning may interfere with the temporally extended processing and encoding of objects at currently attended locations. The amplification of visual processing persists for about 150 ms once an attentional episode is triggered at a particular location may counteract such adverse effects of spatial remapping on ongoing attentional processing, effectively resulting in two simultaneously active attentional foci (see Golomb, 2019, for such a 'dual-spotlight' account of attentional remapping). Evidence for the existence of such an 'attentional trace' at a previously attended location that persists even after a saccade to a new location is executed comes from studies by Golomb and colleagues (Dowd & Golomb, 2020; Golomb et al., 2010b; Golomb et al., 2010a; see also Talsma, White, Mathôt, Munoz, & Theeuwes, 2013). These studies suggest that, for a short period of time, a 'retinotopic attentional trace' lingers after the eye movement. As a result, features that appear in the initial retinotopically mapped location can be erroneously bound with features that appear in the new attended location (Dowd & Golomb, 2019; Golomb, L'Heureux & Kanwisher, 2014).

Summary: The past, present and future of research on selective attention

According to the standard account of selective attention, popularized by Broadbent (1958), attention operates in a temporally discrete fashion at a specific processing stage (after sensory encoding) to select objects for access to higher-level cognitive operations. In this paper, we built upon the insights of "diachronic" computational models (e.g., Reeves & Sperling, 1986; Wyble et al., 2009) and presented the case for an alternative unified diachronic account of attentional selectivity (UDAS). According to this account, selective modulations already emerge during the initial feed-forward stage of perceptual processing. These modulations eventually result in temporally extended periods of intense amplification (attentional episodes) that unfold in tandem with perceptual processing. During these episodes, some stimuli gain access to WM while others fail to do so.

This framework provides a general account of attentional object selection mechanisms whilst avoiding many of the conceptual problems that have haunted attention research for more than half a century. First, it resolves the ambiguity of whether attention should be considered as an effect or as a causal agent that determines that fate of perceptual inputs (Johnston & Dark, 1986). Observers can voluntarily adopt specific task settings (attentional templates), but once these templates are activated, attentional episodes are triggered without further involvement of top-down control, once the presence of a template-matching feature or object is detected. Second, and in line with current neuroscientific understanding, our diachronic framework rejects the notion of a unidirectional chain of perceptual processing and adopts a recurrent processing architecture. This removes the problematic assumption that the function of attentional selection is to protect subsequent capacity-limited processing stages (Allport, 1993). As a result, the diachronic framework removes any fundamental division between pre-attentive and post-attentive processing with a temporally discrete selection process at the interface. By abandoning this traditional dichotomy that is at least implicitly retained in most current models of selective attention, our account provides a new resolution to decades-long early versus late selection debate (see Lachter, Forster, & Ruthruff, 2004, for review).

Looking back: A different perspective on an old problem

Questions about the locus of attentional selection naturally arises when selection is conceived as a discrete process that operates at a specific point within a serial feedforward chain of processing stages. In such standard models, the question when selection occurs is critical, since it determines which

stages operate pre-attentively, and which depend on attention. According to the early-selection view (e.g., Broadbent, 1958; Treisman & Gelade, 1980), attentional selection must occur before information can be fully perceptually processed and identified. Simple visual features can be registered pre-attentively (during "early" parallel sensory processing), whereas feature binding resulting in object representations and the extraction of semantic content (e.g., object identification, word meaning) can only operate after selection has taken place (during 'late' and presumably serial perceptual processing). In contrast, late-selection accounts assume that processes like object identification can occur pre-attentively, and that selection mechanisms therefore operate at subsequent post-perceptual stages involved in memory storage and response selection (e.g., Deutsch & Deutsch, 1963; Duncan, 1980). To identify the locus of selection, two types of evidence were considered to be essential. Observing effects of attentional manipulations on putatively early perceptual processes was interpreted as demonstrating 'early' selection, whereas evidence for the semantic processing of unattended information was seen as supporting 'late' selection accounts. Intense research over several decades produced ample support for both sides of this debate. This led to the idea that instead of always operating at a single fixed locus, selection can be either early or late, depending on the demands of a specific task. For example, Lavie (1995) suggested that this locus is determined by perceptual load. In low-load tasks where target objects can be detected relatively easily, sufficient attentional resources remain available for the identification of unattended distractors. In high-load tasks, all resources are allocated to target-related processing, and unattended distractors are filtered out early, before they are identified. This perceptual load account is appealing because it can explain why evidence for early or late selection can coexist. However, apart from its reliance on the problematic resource concept, it is essentially still a traditional model, because it implies that in any given task context, selection operates at a specific moment in time.

We suggest that the diachronic view can provide a fundamentally different solution to the early versus late selection debate. While previous diachronic models have already gone beyond a strict early\late selection dichotomy (e.g., Bowman & Wyble, 2007; Smith & Ratcliff, 2009), the diachronic framework described here challenges the central assumptions at the core of this dispute. Specifically, it rejects the idea that there is a fundamental qualitative distinction between pre-attentive (early) and post-attentive (late) visual processes. Spatial, temporal, and feature-based attention already modulate the early parallel sensory processing of incoming information (evidence accumulation), and these effects are amplified once an attentional episode is triggered. This is obviously inconsistent with late selection accounts, which assume that attentional selectivity only operates at

post-perceptual processing stages. However, and in contrast to early selection models, parallel evidence accumulation processes extract information not only about low-level features such as colour or shape, but also about higher-level (semantic) attributes such as object identity or category. Furthermore, both featural and semantic representations are subject to the same amplification processes during attentional episodes. There may still be a quantitative difference between processing of low-level features and semantic identities. Evidence accumulation for low-level features starts earlier and/or progresses more rapidly than for semantic identities, resulting in stronger activation levels for feature representations. However, this account rejects a qualitative distinction that assigns the processing of low-level and higher-level attributes to different pre-attentive versus post-attentive stages.

Thus, evidence for or against the semantic processing of unattended information should not be interpreted for hypothetical 'early' or 'late' attentional selection processes. Generally, the degree to which semantic representations affect performance depends on the interplay between attentional enhancement and perceptual competition from other items in the visual field (Benoni & Tsal, 2012). More specifically, the amount of semantic interference produced by distractors depends on whether they are processed within an attentional episode. The identity of entirely unattended distractors can still be registered and affect performance, but these effects are amplified by attention. This was demonstrated in RSVP experiments where both low-level features (Tang et al., 2020) and semantic categories (Marti & Dehaene, 2017) of unattended distractor objects could be successfully decoded from EEG and MEG data for a short duration after stimulus onset. However, the period during which decoding remained above chance was considerably longer for target objects, indicating that the activation levels of both featural and semantic representations were amplified during the attentional episode triggered by these targets. Corresponding behavioural results were found in spatial cueing studies (Remington & Folk, 2001; Zivony & Lamy, 2016a), where the identity of distractors at uncued (unattended) locations elicited reliable compatibility effects, but these effects were considerably larger for cued (attended) distractors.

Although interest in the traditional early versus late selection debate has waned in recent years, this debate has never been truly resolved. Implicit adherence to either of these two alternative accounts may also depend on which aspect of attention is being investigated. While researchers who study spatial attention and visual search often express views in line with early selection (e.g., Luck et al., 2021; Treisman, 2014; Wolfe, 2014, 2021), accounts of attentional selectivity in the temporal domain based on results of RSVP studies almost unanimously adopt a late-selection framework (e.g., Chun & Potter, 1995; Raffone, Srinivasan, &

van Leeuwen, 2014; see Zivony & Lamy, 2021, for review). Such conceptual preferences reflect the enduring influence of the standard temporally discrete concept of selection. We believe that replacing this concept with the alternative diachronic account proposed here can provide a fresh new outlook on the mechanisms of attentional selectivity, and move the field beyond the traditional outdated early/late selection dichotomy.

Looking ahead: Prospects and challenges for the diachronic approach to attentional selectivity

In this paper, we reviewed studies that provide evidence for the existence of attentional episodes, identify the conditions for triggering such episodes (e.g., Goller et al., 2020; Maxwell et al., 2020; Zivony & Lamy, 2018), and delineate the factors that affect their onset latency and duration (Brisson et al., 2007; Callahan-Flintoft & Wyble, 2017; Foster et al., 2020; Töllner et al., 2011; Zivony & Eimer, 2021a). Many other aspects of the attentional blink phenomena that we could not discuss here are also in line with our general framework. For example, current research strongly supports the idea that the costs observed during the attentional blink period is caused by the suppression of attentional engagement (Nieuwenstein et al., 2005; Olivers & Meeter, 2008; Wyble et al., 2009, 2011; Zivony & Lamy, 2016b; see Zivony & Lamy, 2021, for review). Our diachronic framework of attentional selectivity, and specifically the concept of attentional episodes, has the potential to provide a unified account of a wide variety of findings in the attention literature, such as contingent capture cuing effects, distractor compatibility effects, lag-1 sparing, distractor intrusions, order reversals, temporal integration, and object-substitution masking.

As systematic research into attentional episodes has only recently gathered momentum, many questions about their basic properties and about their role in the attentional control of different perceptual and cognitive processes still remain to be answered. The three main parameters of attentional episodes are their duration, their amplitude (i.e., the amount of processing amplification afforded to visual stimuli within an episode), and the engagement threshold for triggering an episode. Given the wide variety of different contexts in which attentional selectivity is required, it appears unlikely that these three parameters are always set at a single constant value. Task-dependent flexibility in these parameters should be highly adaptive. For example, changes to the engagement threshold may allow episodes to be triggered faster under conditions where target objects are highly likely to appear at particular locations, time points, or to have a specific feature. Changing the amplitude of an episode can facilitate the processing of prioritized objects, and changes to its duration could be a way to regulate how many objects

will gain access to WM. Investigating whether and how task manipulations affect the properties of attentional episodes will be challenging because any effects on these episodes must be differentiated from changes to earlier selective mechanisms (such as spatial attention, temporal attention or feature-based attention). This will require the development of new experimental designs and possibly the identification of distinct neural markers for the three defining parameters of attentional episodes.

Previous studies have already addressed the duration of attentional episodes, and the factors that can determine their length, although so far without any firm conclusions. It has been suggested that episodes can be as short as 150 ms (Olivers & Meeter, 2008; Wyble et al., 2011), or may extend to up to a second (Reeves & Sperling, 1986). One factor that appears to affect this duration is whether the object that triggers an episode is followed by additional targets or by task-irrelevant objects (Olivers et al., 2007; Olivers & Meeter, 2008; but see also Callahan-Flintoft, Chen, & Wyble, 2018; Tan & Wyble, 2015). While some studies have suggested that processing amplification can be maintained for up to about four consecutive relevant items (Kawahara et al., 2006; Olivers et al., 2007), there are performance costs relative to single-target processing (Dell'Acqua et al., 2012), which could be due to perceptual competition between multiple representations within a single episode, competitive interactions in WM, or an overall reduction in attentional amplification. These and other open questions illustrate that improved insights into the temporal dynamics of attentional episodes is clearly required, also because this would provide a crucial link between artificial lab tasks and real-life tasks that require continuous intake of changing information (e.g., reading). More generally, adopting a diachronic perspective on selectivity in visual processing can generate novel questions that will also have an impact on translational and applied research.

Finally, the diachronic framework can provide a critical step towards a coherent conceptualization of attentional selectivity, which is essential for future progress in attention research. While it is undoubtedly correct that terms such as 'attention' and 'selection' have been poorly defined and often been used inconsistently in the past, we do not share the pessimistic view that these concepts are generally unsuited for scientific inquiry (Anderson, 2011; Di Lollo, 2018; Hommel et al., 2019). Rather than abandoning any reference to selective attention altogether, we suggest that we need to sharpen these conceptual tools in order to avoid logical pitfalls, such as defining attention both as the selector of information and the process that enhances the selected information. While we focused on attentional episodes as a core mechanism for the selection of task-relevant information, the diachronic approach outlined here emphasizes that all attentional processes unfold over time. Serious consideration of this temporal dimension will help to dissolve many

of the confusions and controversies of the past. The future progress of attention research as a mature field of scientific inquiry may depend on it.

References

Akyürek, E. G., & Meijerink, S. K. (2012). The deployment of visual attention during temporal integration: An electrophysiological investigation. *Psychophysiology, 49*(7), 885-898.

Akyürek, E. G., & Wolff, M. J. (2016). Extended temporal integration in rapid serial visual presentation: Attentional control at Lag 1 and beyond. *Acta Psychologica, 168*, 50-64.

Akyürek, E. G., Eshuis, S. A., Nieuwenstein, M. R., Saija, J. D., Başkent, D., & Hommel, B. (2012). Temporal target integration underlies performance at lag 1 in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 38*(6), 1448-1464.

Akyürek, E. G., Kappelmann, N., Volkert, M., & Van Rijn, H. (2017). What you see is what you remember: Visual chunking by temporal integration enhances working memory. *Journal of Cognitive Neuroscience, 29*(12), 2025–2036.

Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In: *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience*, (pp 183-218). MIT Press.

Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology*, *2*, 246.

Anderson, B. A. (2016). The attention habit: How reward learning shapes attentional selection. *Annals of the New York Academy of Sciences*, *1369*(1), 24-39.

Anderson, B. A. (2021). Time to stop calling it attentional "capture" and embrace a mechanistic understanding of attentional priority. *Visual Cognition, 29*(9), 537–540. https://doi.org/10.1080/13506285.2021.1892894.

Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In: *Advances in psychology* (Vol. 99, pp. 369-399). North-Holland.

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437-443.

Bachman, M. D., Wang, L., Gamble, M. L., & Woldorff, M. G. (2020). Physical salience and value-driven salience operate through different neural mechanisms to enhance attentional selection. *Journal of Neuroscience, 40*(28), 5455-5464.

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences*, *34*(4), 210-224.

Benoni, H., & Tsal, Y. (2012). Controlling for dilution while manipulating load: perceptual and sensory limitations are just two aspects of task difficulty. *Psychonomic Bulletin & Review, 19*(4), 631-638.

Boehler, C. N., Schoenfeld, M. A., Heinze, H. J., & Hopf, J. M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences, 105*(25), 8742-8747.

Botella, J., Suero, M., & Barriopedro, M. I. (2001). A model of the formation of illusory conjunctions in the time domain. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1452–1467.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review*, *114*(1), 38–70.

Brisson, B., Robitaille, N., & Jolicoeur, P. (2007). Stimulus intensity affects the latency but not the amplitude of the N2pc. *Neuroreport*, *18*(15), 1627-1630.

Broadbent, D. E. (1958). *Perception and communication*. Oxford University Press.

Brockmole, J. R., Wang, R. F., & Irwin, D. E. (2002). Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 315-334.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological Review*, *112*(2), 291–328.

Callahan-Flintoft, C., & Wyble, B. (2017). Non-singleton colours are not attended faster than categories, but they are encoded faster: A combined approach of behavior, modeling and ERPs. *Vision Research*, *140*, 106-119.

Callahan-Flintoft, C., Chen, H., & Wyble, B. (2018). A hierarchical model of visual processing simulates neural mechanisms underlying reflexive attention. *Journal of Experimental Psychology: General, 147*(9), 1273-1294.

Camprodon, J. A., Zohary, E., Brodbeck, V., & Pascual-Leone, A. (2010). Two phases of V1 activity for visual recognition of natural images. *Journal of Cognitive Neuroscience*, *22*(6), 1262-1269.

Carmel, T., & Lamy, D. (2014). The same-location cost is unrelated to attentional settings: An object-updating account. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1465-1478.

Chambers, C. D., & Mattingley, J. B. (2005). Neurodisruption of selective attention: insights and implications. *Trends in Cognitive Sciences*, *9*(11), 542-550.

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.

Chen, H., & Wyble, B. (2018). The neglected contribution of memory encoding in spatial cueing: A new theory of costs and benefits. *Psychological Review, 125*(6), 936-968.

Chen, J., Leber, A. B., & Golomb, J. D. (2019). Attentional capture alters feature perception. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(11), 1443–1454.

Christie, G. J., Livingstone, A. C., & McDonald, J. J. (2015). Searching for inefficiency in visual search. *Journal of Cognitive Neuroscience*, *27*(1), 46-56.

Chun, M. M. (1997). Temporal binding errors are redistributed by the attentional blink. *Perception & Psychophysics, 59*, 1191–1199.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360-365.

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), 109-127.

Cohen, M. R., & Maunsell, J. H. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience, 12*(12), 1594-1600.

Cohen, J. Y., Heitz, R. P., Schall, J. D., & Woodman, G. F. (2009). On the origin of event-related potentials indexing covert attentional selection during visual search. *Journal of Neurophysiology*, *102*(4), 2375-2386.

Darnell, M., & Lamy, D. (2021). Spatial cueing effects do not always index attentional capture: Evidence for a Priority Accumulation Framework. *Psychological Research*. https://doi.org/10.1007/s00426-021-01597-0.

Dell'Acqua, R., Dux, P. E., Wyble, B., & Jolicoeur, P. (2012). Sparing from the attentional blink is not spared from structural limitations. *Psychonomic Bulletin & Review, 19*(2), 232-238.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193-222.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*(1), 80-90.

Di Lollo, V. (2014). Reentrant processing mediates object substitution masking: Comment on Põder (2013). *Frontiers in Psychology, 5*, 819.

Di Lollo, V. (2018). Attention is a sterile concept; iterative reentry is a fertile substitute. *Consciousness and Cognition, 64*, 45-49.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*(4), 481-507.

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

Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping, 15*(2), 95-111.

Dowd, E. W., & Golomb, J. D. (2019). Object-feature binding survives dynamic shifts of spatial attention. *Psychological Science, 30*(3), 343-361.

Dowd, E. W., & Golomb, J. D. (2020). The Binding Problem after an eye movement. *Attention, Perception, & Psychophysics, 82*(1), 168-180.

Drisdelle, B. L., & Jolicoeur, P. (2018). Early and late selection processes have separable influences on the neural substrates of attention. *International Journal of Psychophysiology, 127*, 52-61.

Drisdelle, B. L., West, G. L., & Jolicoeur, P. (2016). The deployment of visual spatial attention during visual search predicts response time: electrophysiological evidence from the N2pc. *Neuroreport, 27*(16), 1237-1242.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology, 92*(1), 53-78.

Driver, J., & Frith, C. (2000). Shifting baselines in attention research. *Nature Reviews Neuroscience, 1*(2), 147-148.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*(5040), 90-92.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87*(3), 272-300.

Duncan, J., & Humphreys, G. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance, 18*(2), 578–588.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and clinical neurophysiology, 99*(3), 225-234.

Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 20*(8), 1423-1433.

Eimer, M., & Mazza, V. (2005). Electrophysiological correlates of change detection. *Psychophysiology, 42*(3), 328-342.

Eimer, M., Kiss, M., & Cheung, T. (2010). Priming of pop-out modulates attentional target selection in visual search: Behavioural and electrophysiological evidence. *Vision Research, 50*(14), 1353-1361.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143-149.

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience, 25*(5), 754-761.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *Journal of Vision, 8*(1):12, 1-12.

Failing, M., & Theeuwes, J. (2018). Selection history: How reward modulates selectivity of visual attention. *Psychonomic Bulletin & Review, 25*(2), 514-538.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 847-858.

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), 1030-1044.

Foster, J. J., Bsales, E. M., & Awh, E. (2020). Covert spatial attention speeds target individuation. *Journal of Neuroscience, 40*(13), 2717-2726.

Gaspelin, N., Ruthruff, E., & Lien, M. C. (2016). The problem of latent attentional capture: Easy visual search conceals capture by task-irrelevant abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 42*(8), 1104-1120.

Goller, F., Schoeberl, T., & Ansorge, U. (2020). Testing the top-down contingent capture of attention for abrupt-onset cues: Evidence from cue-elicited N2pc. *Psychophysiology, 57*(11), e13655.

Golomb, J. D. (2019). Remapping locations and features across saccades: a dual-spotlight theory of attentional updating. *Current Opinion in Psychology, 29*, 211-218.

Golomb, J. D., Nguyen-Phuc, A. Y., Mazer, J. A., McCarthy, G., & Chun, M. M. (2010a). Attentional facilitation throughout human visual cortex lingers in retinotopic coordinates after eye movements. *Journal of Neuroscience, 30*(31), 10493-10506.

Golomb, J. D., Pulido, V. Z., Albrecht, A. R., Chun, M. M., & Mazer, J. A. (2010b). Robustness of the retinotopic attentional trace after eye movements. *Journal of Vision, 10*(3):19, 1-12.

Golomb, J. D., L'Heureux, Z. E., & Kanwisher, N. (2014). Feature-binding errors after eye movements and shifts of attention. *Psychological Science, 25*(5), 1067-1078.

Goodbourn, P. T., & Holcombe, A. O. (2015). "Pseudoextinction": Asymmetries in simultaneous attentional selection. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 364–384.

Grubert, A., & Eimer, M. (2018). The time course of target template activation processes during preparation for visual search. *Journal of Neuroscience, 38*(44), 9527-9538.

Grubert, A., & Eimer, M. (2020). Preparatory Template Activation during Search for Alternating Targets. *Journal of Cognitive Neuroscience*, 1-11.

Hackley, S. A., Schankin, A., Wohlschlaeger, A., & Wascher, E. (2007). Localization of temporal preparation effects via trisected reaction time. *Psychophysiology, 44*(2), 334-338.

Hayden, B. Y., & Gallant, J. L. (2005). Time course of attention reveals different mechanisms for spatial and feature-based attention in area V4. *Neuron, 47*(5), 637-643.

Hilkenmeier, F., Olivers, C. N., & Scharlau, I. (2012a). Prior entry and temporal attention: Cueing affects order errors in RSVP. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 180-190.

Hilkenmeier, F., Scharlau, I., Weiß, K., & Olivers, C. N. (2012b). The dynamics of prior entry in serial visual processing. *Visual Cognition, 20*(1), 48-76.

Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into

a single episode. *The Quarterly Journal of Experimental Psychology Section A, 58*(8), 1415-1433.

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics, 81*(7), 2288-2303.

Hopf, J. M., Boelmans, K., Schoenfeld, M. A., Luck, S. J., & Heinze, H. J. (2004). Attention to features precedes attention to locations in visual search: evidence from electromagnetic brain responses in humans. *Journal of Neuroscience, 24*(8), 1822-1832.

Hopf, J. M., Luck, S. J., Boelmans, K., Schoenfeld, M. A., Boehler, C. N., Rieger, J., & Heinze, H. J. (2006). The neural site of attention matches the spatial scale of perception. *Journal of Neuroscience, 26*(13), 3532-3540.

Huber-Huber, C., Ditye, T., Fernández, M. M., & Ansorge, U. (2016). Using temporally aligned event-related potentials for the investigation of attention shifts prior to and during saccades. *Neuropsychologia, 92*, 129-141.

Ibos, G., Duhamel, J. R., & Hamed, S. B. (2013). A functional hierarchy within the parietofrontal network in stimulus selection and attention control. *Journal of Neuroscience, 33*(19), 8359-8369.

Itthipuripat, S., Garcia, J. O., Rungratsameetaweemana, N., Sprague, T. C., & Serences, J. T. (2014). Changing the spatial scope of attention alters patterns of neural gain in human cortex. *Journal of Neuroscience, 34*(1), 112-123.

Itthipuripat, S., Deering, S., & Serences, J. T. (2019). When conflict cannot be avoided: Relative contributions of early selection and frontal executive control in mitigating Stroop conflict. *Cerebral Cortex, 29*(12), 5037-5048.

James, W (1890). *Principles of psychology*. Holt.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology, 37*(1), 43-75.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Reviews Neuroscience, 1*(2), 91-100.

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: different processes and overlapping neural systems. *The Neuroscientist, 20*(5), 509-521.

Kawahara, J. I., Kumada, T., & Di Lollo, V. (2006). The attentional blink is governed by a temporary loss of control. *Psychonomic Bulletin & Review, 13*(5), 886-890.

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences, 6*(3), 120-125.

Kiss, M., Van Velzen, J., & Eimer, M. (2008). The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology, 45*(2), 240-249.

Kiss, M., Driver, J., & Eimer, M. (2009). Reward priority of visual target singletons modulates event-related potential signatures of attentional selection. *Psychological Science, 20*(2), 245-251.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): still no identification without attention. *Psychological Review, 111*(4), 880-913.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*(11), 571-579.

Lamy, D., Leber, A., & Egeth, H. E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology: Human Perception and Performance, 30*(6), 1019-1031.

Lamy, D., Darnell, M., Levi, A., & Bublil, C. (2018). Testing the attentional dwelling hypothesis of attentional capture. *Journal of Cognition, 1*(1), 43.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 451-468.

Lennie, P. (2003). The cost of cortical computation. *Current Biology 13*(6): 493–497.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997a). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77*(1), 24-42.

Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997b). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology, 33*(1), 64-87.

Luck, S. J., Gaspelin, N., Folk, C. L., Remington, R. W., & Theeuwes, J. (2021). Progress toward resolving the attentional capture debate. *Visual Cognition, 29*(1), 1-21.

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews, 62*, 100-108.

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition, 22*(6), 657-672.

Marti, S., & Dehaene, S. (2017). Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nature Communications, 8*(1), 1-13.

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology, 14*(9), 744-751.

Maxwell, J. W., Gaspelin, N., & Ruthruff, E. (2020). No identification of abrupt onsets that capture attention: evidence against a unified model of spatial attention. *Psychological Research.* https://doi.org/10.1007/s00426-020-01367-4.

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

Moore, T., & Fallah, M. (2004). Microstimulation of the frontal eye field and its effects on covert spatial attention. *Journal of Neurophysiology, 91*(1), 152-162.

Müller, H. J., & Findlay, J. M. (1988). The effect of visual attention of peripheral discrimination thresholds in single and multiple element displays. *Acta Psychologica, 69*(2), 129-155.

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

Neisser, U. (1967). *Cognitive Psychology*. Appleton-Century-Crofts.

Nieuwenstein, M. R. (2006). Top-down controlled, delayed selection in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 973-985.

Nieuwenstein, M. R., Chun, M. M., van der Lubbe, R. H., & Hooge, I. T. (2005). Delayed attentional engagement in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1463-1475.

Noudoost, B., Chang, M. H., Steinmetz, N. A., & Moore, T. (2010). Top-down control of visual attention. *Current Opinion in Neurobiology, 20*(2), 183-190.

Oberauer, K., & Lin, H. Y. (2017). An interference model of visual working memory. *Psychological Review, 124*(1), 21-59.

Ogawa, T., & Komatsu, H. (2006). Neuronal dynamics of bottom-up and top-down processes in area V4 of macaque monkeys performing a visual search. *Experimental Brain Research, 173*(1), 1-13.

Olivers, C. N., & Meeter, M. (2008). A boost and bounce theory of temporal attention. *Psychological Review, 115*(4), 836–863.

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*, 126–139.

Olivers, C. N., Hilkenmeier, F., & Scharlau, I. (2011). Prior entry explains order reversals in the attentional blink. *Attention, Perception, & Psychophysics, 73*(1), 53-67.

Perrone-Bertolotti, M., Tiali, S. E. B., Vidal, J. R., Petton, M., Croize, A. C., Deman, P., Minotti, L., Bhattacharjee, M., Baciu, M., Kahane, P. & Lachaux, J. P. (2020). A real-time marker of object-based attention in the human brain. A possible component of a "gate-keeping mechanism" performing late attentional selection in the Ventro-Lateral Prefrontal Cortex. *NeuroImage*, *210*, 116574.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73-89.

Põder, E. (2013). Attentional gating models of object substitution. *Journal of Experimental Psychology: General, 142*(4), 1130-1141. https://doi.org/10.1037/a0030575

Pooresmaeili, A., Poort, J., & Roelfsema, P. R. (2014). Simultaneous selection by object-based attention in visual and frontal cortex. *Proceedings of the National Academy of Sciences*, *111*(17), 6467-6472.

Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: the visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 979-992.

Potter, M. C., Staub, A., & O'Connor, D. H. (2002). The time course of competition for attention: Attention is initially labile. *Journal of Experimental Psychology: Human Perception and Performance, 28*(5), 1149-1162.

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, *76*(2), 270-279.

Pratt, J., Hillis, J., & Gold, J. M. (2001). The effect of the physical characteristics of cues and targets on facilitation and inhibition. *Psychonomic Bulletin & Review, 8*(3), 489-495.

Purcell, B. A., Schall, J. D., & Woodman, G. F. (2013). On the origin of event-related potentials indexing covert attentional selection during visual search: timing of selection by macaque frontal eye field and event-related potentials during pop-out search. *Journal of Neurophysiology*, *109*(2), 557-569.

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22*(8), 1336-1344.

Raffone, A., Srinivasan, N., & van Leeuwen, C. (2014). The interplay of attention and consciousness in visual search, attentional blink and working memory consolidation. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences, 369*, 20130215.

Ramgir, A., & Lamy, D. (2021). Does feature intertrial priming guide attention? The jury is still out. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01997-8

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260-281.

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), 849–860.

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

Remington, R. W., & Folk, C. L. (2001). A dissociation between attention and selection. *Psychological Science*, *12*(6), 511-515.

Salahub, C. M., & Emrich, S. M. (2018). ERP evidence for temporal independence of set size and object updating in object substitution masking. *Attention, Perception, & Psychophysics*, *80*(2), 387-401.

Salthouse, T. A., & Ellis, C. L. (1980). Determinants of eye-fixation duration. *The American Journal of Psychology*, 207-234.

Sani, I., Santandrea, E., Morrone, M. C., & Chelazzi, L. (2017). Temporally evolving gain mechanisms of attention in macaque area V4. *Journal of Neurophysiology*, *118*(2), 964-985.

Seibold, V. C., Stepper, M. Y., & Rolke, B. (2020). Temporal attention boosts perceptual effects of spatial attention and feature-based attention. *Brain and Cognition, 142*, 105570.

Seth, A. K., & Baars, B. J. (2005). Neural Darwinism and consciousness. *Consciousness and Cognition*, *14*(1), 140-168.

Shih, S. I. (2008). The attention cascade model and attentional blink. *Cognitive Psychology*, *56*(3), 210-236.

Shih, S. I., & Reeves, A. (2007). Attentional capture in rapid serial visual presentation. *Spatial Vision*, *20*(4), 301-315.

Simione, L., Akyürek, E. G., Vastola, V., Raffone, A., & Bowman, H. (2017). Illusions of integration are subjectively impenetrable: Phenomenological experience of Lag 1 percepts during dual-target RSVP. *Consciousness and Cognition, 51*, 181-192.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2008). Are there multiple visual short-term memory stores?. *PLOS One, 3*(2), e1699.

Slotnick, S. D. (2018). The experimental parameters that affect attentional modulation of the ERP C1 component. *Cognitive Neuroscience*, *9*(1-2), 53-62.

Smith, P. L., & Wolfgang, B. J. (2004). The attentional dynamics of masked detection. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 119-136.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological review, 116*(2), 283–317.

Spalek, T. M., Lagroix, H. E., Yanko, M. R., & Di Lollo, V. (2012). Perception of temporal order is impaired during the time course of the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 402-413.

Spence, C., & Parise, C. (2010). Prior-entry: A review. *Consciousness and Cognition, 19*(1), 364-379.

Taatgen, N. A., Juvina, I., Schipper, M., Borst, J. P., & Martens, S. (2009). Too much control can hurt: A threaded cognition model of the attentional blink. *Cognitive Psychology, 59*, 1–29.

Talcott, T. N., & Gaspelin, N. (2021). Eye movements are not mandatorily preceded by the N2pc component. *Psychophysiology*, e13821. https://doi.org/10.1111/psyp.13821

Talsma, D., White, B. J., Mathôt, S., Munoz, D. P., & Theeuwes, J. (2013). A retinotopic attentional trace after saccadic eye movements: evidence from event-related potentials. *Journal of Cognitive Neuroscience, 25*(9), 1563–1577.

Tamber-Rosenau, B. J., & Marois, R. (2016). Central attention is serial, but midlevel and peripheral attention are parallel—A hypothesis. *Attention, Perception, & Psychophysics*, *78*(7), 1874-1888.

Tan, M., & Wyble, B. (2015). Understanding how visual attention locks on to a location: Toward a computational model of the N 2pc component. *Psychophysiology*, *52*(2), 199-213.

Tang, M. F., Ford, L., Arabzadeh, E., Enns, J. T., Visser, T. A., & Mattingley, J. B. (2020). Neural dynamics of the attentional blink revealed by encoding orientation selectivity during rapid visual presentation. *Nature Communications, 11*(1), 1-14.

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica*, *135*(2), 77-99.

Titchener, E. B. (1928). *A textbook of Psychology*. New-York: The Macmillan Company.

Töllner, T., Zehetleitner, M., Gramann, K., & Müller, H. J. (2011). Stimulus saliency modulates pre-attentive processing speed in human visual cortex. *PLoS One, 6*(1), e16276.

Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The retinotopy of visual spatial attention. *Neuron*, *21*(6), 1409-1422.

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

Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1373), 1295-1306.

Treisman, A. (2014). The psychological reality of levels of processing. *Levels of Processing in Human Memory*, 301–330.

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

Vernet, M., Quentin, R., Chanes, L., Mitsumasu, A., & Valero-Cabré, A. (2014). Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. *Frontiers in Integrative Neuroscience*, *8*(66), 1-24.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748-751.

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

Vul, E., Hanus, D., & Kanwisher, N. (2009). Attention as inference: Selection is probabilistic; responses are all-or-none samples. *Journal of Experimental Psychology: General, 138*, 546–560.

Warren, S. G., Yacoub, E., & Ghose, G. M. (2014). Featural and temporal attention selectively enhance task-appropriate representations in human primary visual cortex. *Nature Communications*, *5*(1), 1-12.

Weaver, M. D., van Zoest, W., & Hickey, C. (2017). A temporal dependency account of attentional inhibition in oculomotor control. *NeuroImage, 147*, 880–894.

Westerberg, J. A., Schall, M. S., Maier, A., Woodman, G. F., & Schall, J. D. (2021). Laminar microcircuitry of visual cortex producing attention-associated electric fields. *bioRxiv*. https://doi.org/10.1101/2021.09.09.459639.

Wilschut, A., Theeuwes, J., & Olivers, C. N. (2011). The time course of attention: selection is transient. *PLoS One, 6*(11), e27661.

Wilschut, A., Theeuwes, J., & Olivers, C. N. (2013). Early perceptual interactions shape the time course of cueing. *Acta Psychologica, 144*(1), 40-50.

Wilschut, A., Theeuwes, J., & Olivers, C. N. (2015). Nonspecific competition underlies transient attention. *Psychological Research, 79*(5), 844-860.

Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided search. *The Oxford Handbook of Attention*, *11*, 35–44.

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01859-9.

Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*(6747), 867-869.

Woodman, G. F., & Luck, S. J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychological Science, 14*(6), 605-611.

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.

Wyble, B., Potter, M. C., Bowman, H., & Nieuwenstein, M. (2011). Attentional episodes in visual perception. *Journal of Experimental Psychology: General*, *140*(3), 488–505.

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12*(1), 24-25.

Zhou, H., & Desimone, R. (2011). Feature-based attention in the frontal eye field and area V4 during visual search. *Neuron, 70*(6), 1205-1217.

Zhou, H., Schafer, R. J., & Desimone, R. (2016). Pulvinar-cortex interactions in vision and attention. *Neuron*, *89*(1), 209-220.

Zivony, A. & Eimer, M. (2020). Perceptual competition between targets and distractors determines working memory access and produces intrusion errors in RSVP tasks. *Journal of Experimental Psychology: Human Perception & Performance*. 46(12), 1490–1510.

Zivony, A. & Eimer, M. (2021a). Distractor intrusions are the result of delayed attentional engagement: a new temporal variability account of attentional selectivity in dynamic visual tasks. *Journal of Experimental Psychology: General*, *150*(1), 23–41.

Zivony, A. & Eimer, M. (2021b). The number of expected targets modulates access to working memory: a new unified account of lag-1 sparing and distractor intrusions. *Journal of Experimental Psychology: Human Perception & Performance*. https://doi.org/10.1037/xhp0000958.

Zivony, A. & Lamy, D. (2016a). The role of dimension relevance in features' access to response-selection mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 42*(11), 1873-1885.

Zivony, A. & Lamy, D. (2016b). Attentional capture and engagement during the attentional blink: a "camera" metaphor of attention. *Journal of Experimental Psychology: Human Perception and Performance, 42*(11), 1886–1902.

Zivony, A. & Lamy, D. (2018). Contingent attentional engagement: stimulus- and goal-driven capture have qualitatively different consequences. *Psychological Science, 29*, 1930-1941.

Zivony, A. & Lamy, D. (2021). What processes are disrupted during the attentional blink? An integrative review of event-related potential research. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01973-2.

Zivony, A., Allon, A. S., Luria, R. & Lamy, D. (2018). Dissociating between the N2pc and attentional shifting: an attentional blink study. *Neuropsychologia*, 121, 153-163.