Abbreviated title: NORMALIZATION MODEL OF TEMPORAL ATTENTION

A dynamic normalization model of temporal attention

Rachel N. Denison, Marisa Carrasco, David J. Heeger

Department of Psychology and Center for Neural Science, New York University, New York, NY 10003, USA

Corresponding author:
Rachel N. Denison
Department of Psychology, New York University
6 Washington Place
New York, NY 10003
rachel.denison@nyu.edu

Author contributions: RND, MC, and DJH conceived of the project and designed the experiment. RND and DJH conceived of the model. RND implemented the model, conducted the experiment, and analyzed the data. RND wrote and all authors edited the manuscript.

Competing interests: The authors have no competing interests to declare.

Acknowledgements: This research was supported by National Institutes of Health National Eye Institute R01 EY019693 to MC and DJH and R01 EY016200 to MC, F32 EY025533 to RND, and T32 EY007136 to NYU. We thank Hsin-Hung Li for consultation on the model, Valentina Peña for assistance with data collection, and Antonio Fernández and Michael Jigo for their comments on the manuscript.

Abstract

Vision is dynamic, handling a continuously changing stream of input, yet most models of visual attention are static. Here, we develop a dynamic normalization model of visual temporal attention and constrain it with new psychophysical human data. We manipulated temporal attention–the prioritization of visual information at specific points in time–to a sequence of two stimuli separated by a variable time interval. Voluntary temporal attention improved perceptual sensitivity only over a specific interval range. To explain these data, we modeled voluntary and involuntary attentional gain dynamics. Voluntary gain enhancement took the form of a limited resource over short time intervals, which recovered over time. Taken together, our theoretical and experimental results formalize and generalize the idea of limited attentional resources across space at a single moment to limited resources across time at a single location.

Introduction

The visual system receives continuous, time-varying input, which must be dynamically prioritized according to behavioral goals. However, most data and theory on visual perception and attention have been motivated by a static picture of visual processing, focusing on how we see a single image that is isolated in time. Here we generalized a successful static computational theory of visual and attentional processing into a dynamic model, which we constrained using our novel psychophysical protocol and new data on the dynamics of temporal attention.

Our theory is based on the principle of normalization. The normalization model explains contextual modulation in neural populations with divisive suppression [1, 2]. Normalization appears to be widespread in both basic sensory [2-6] and higher-order perceptual and cognitive processing [7-12]. For this reason, it has been described as a "canonical cortical computation" [1].

Several models of attention combine sensory normalization with attentional modulation [6, 12-18]. In these models, attention changes the sensitivity of neural responses to sensory inputs by modulating the gain of the responses. One such model, developed by Reynolds and Heeger [12], proposes that attention modulates neural activity before normalization. This formulation has reconciled ostensibly conflicting electrophysiological and psychophysical findings [12] and predicted new results that have been empirically confirmed [19-21]. However, this leading theory of spatial and feature-based attention is static, with no time component.

Dynamic normalization models have been developed to account for the time courses of neuronal responses [5, 10, 22-25] and dynamic sensory processes like adaptation [2, 26-30]. But these models have not incorporated attention. It has been noted[17] that differential shunting equations can be used to implement normalization, as in shunting equation models of spatial attention [31, 32].

A major challenge in developing a dynamic normalization model of attention is establishing what the attentional gain dynamics actually are. The behavioral time courses of spatial attention have been characterized: voluntary spatial attention takes 300 ms to be allocated, and involuntary spatial attention peaks at 90-120 ms [33-35]. But visual attention is not only directed to locations in space; it is also directed to points in time.

Temporal attention is the prioritization of visual information at specific points in time – for example, the moment a behaviorally relevant stimulus will appear [36]. Even with spatial attention fixed at one location, visual temporal attention can be manipulated using temporal precues to specific time points. Such voluntary, or goal-directed, temporal attention affects perception [37-41], neural responses [41-45], and microsaccades [46]. Voluntary temporal attention can lead to both perceptual benefits at attended times and perceptual costs at unattended times, relative to when attention is distributed across time [38]. But the temporal dynamics of attention that lead to these benefits and costs are unknown. Moreover, there are no existing models of voluntary attention to specific time points.

We define involuntary temporal attention as stimulus-driven attentional dynamics that prioritize specific points in time in a non-goal-directed fashion – for example, an increase in attention following a salient stimulus. Involuntary *spatial* attention transiently enhances visual processing at a stimulated location and its underlying mechanisms are at least partially distinct from those underlying voluntary spatial attention [33, 35]. However, the dynamics of involuntary *temporal* attention (even when spatial attention is fixed) are unknown.

We developed a normalization model of dynamic attention that can capture not only spatial and feature-based attention but also temporal attention. We performed a psychophysical experiment to measure how voluntary and involuntary temporal attention affect perception across time, and we used these new data on temporal attentional dynamics to constrain the model. The model that best fits the data predicts a limitation in the availability of voluntary attentional gain across time intervals of ~1 s.

Results

Behavior

To determine the dynamics of voluntary and involuntary temporal attention, we performed a behavioral experiment (**Figure 1a,b**). Observers judged the orientation of gratings while voluntary temporal attention was directed to different points in time. On each trial, two gratings appeared in sequence at the same location, separated by a stimulus onset asynchrony (SOA). The SOA ranged from 100-800 ms across testing sessions but was fixed within a session to ensure predictable stimulus timing. Voluntary temporal attention was manipulated by an auditory precue to attend the first target (T1), the second target (T2), or both targets (neutral precue). On valid attention trials (60%), observers were asked at the end of the trial to report the orientation of the precued target. On invalid attention trials (20%), they were asked to report the target that was not precued. And on neutral trials (20%), they were equally likely to be asked to report T1 or T2. Therefore, only the time point(s) to which voluntary attention was directed varied from trial to trial.

Critically, this two-target temporal precueing protocol, which we developed in Ref. 38, allowed us to measure how voluntary temporal attention affected the perception of both targets as a function of SOA–which was necessary to infer voluntary attentional dynamics. Such measurement could not have been accomplished using previous temporal precueing protocols, which presented only one target per trial, so observers could reorient attention to the second time point if no target appeared at the first. This protocol also allowed us to investigate involuntary attentional dynamics by assessing the impact of involuntary attention elicited by T1 on T2 behavior, as a function of SOA. Behavioral performance depended on the temporal attentional precue, the SOA, and the reported target. We identified four main features of the behavioral data.

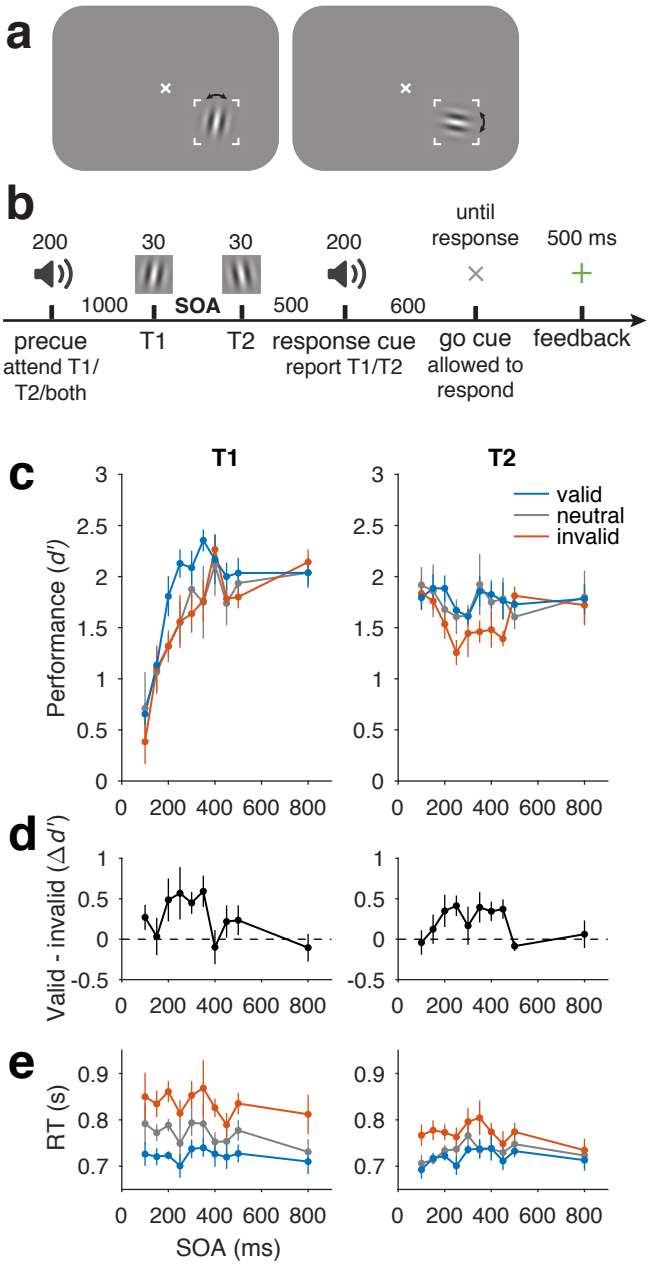

Figure 1. Behavioral protocol and data. **a)** Stimulus display and task schematic. Observers discriminated clockwise (CW) vs. counterclockwise (CCW) tilts about either the vertical (left) or horizontal (right) axis. (Arrows were not shown to observers.) **b)** Trial sequence. Tilts and axes were independent for T1 and T2. SOA varied across testing sessions from 100-800 ms but was fixed within each session. **c)** Perceptual sensitivity (*d'*), for each SOA, precue (valid, neutral, invalid), and target interval (T1, T2). **d)** Effect of precueing temporal attention on perceptual sensitivity (i.e., difference in performance between valid and invalid precues). **e)** Reaction time (RT). Error bars in **c** and **e** are within-observers SEM computed separately for each target to reflect differences among conditions for each target. Error bars in **d** are SEM.

First, voluntary temporal attention affected behavior, resulting in attentional tradeoffs between the two targets. Overall, perceptual sensitivity (*d'*) was highest for valid trials, lowest for invalid trials, and intermediate for neutral trials (**Figure 1c,d**). In a repeated-measures ANOVA with precue validity, SOA, and target as factors, there was a main effect of validity, $F(2, 8) = 8.85$, p

= 0.0094, η_G^2 = 0.048. Temporal precueing tended to produce attentional benefits for T1 (valid better than neutral, which was similar to invalid) but attentional costs for T2 (invalid worse than neutral, which was similar to valid). Planned repeated-measures ANOVAs assessing benefits and costs separately for T1 and T2 showed a trend toward significant benefits for T1, F(1, 4) = 5.00, p = 0.089, η_G^2 = 0.079, but not costs, F(1, 4) = 0.43, p = 0.55, η_G^2 = 0.0017, and significant costs for T2, F(1, 4) = 15.10, p = 0.018, η_G^2 = 0.065, but not benefits, F(1, 4) = 0.089, p = 0.78, η_G^2 = 0.0015. Reaction time showed a similar dependence on the attentional precue, with fastest responses for valid trials, slowest for invalid trials, and intermediate responses for neutral trials (**Figure 1e**; main effect of validity, F(2, 8) = 21.92, p < 0.001, η_G^2 = 0.27), confirming that speed-accuracy tradeoffs did not drive the differences in *d'*. The presence of precueing effects, and particularly the attentional benefits for one target mirrored by attentional costs for the other, indicates temporal attentional tradeoffs across time, consistent with our previous findings with a 250 ms SOA [38].

Second, the temporal precue affected perceptual sensitivity differently at different SOAs. The precue had its largest effects at intermediate SOAs, 200-350 ms for T1 and 200-450 ms for T2, and little or no effect at the shortest and longest SOAs. This SOA dependence can be seen in **Figure 1d**, where we plot the difference between *d'* values for trials with valid and invalid precues. Confirming this observation, a repeated-measures ANOVA of the precueing effect (valid – invalid) with target and SOA as factors showed a main effect of SOA, F(9, 36) = 3.13, p = 0.0069, η_G^2 = 0.20. There was neither a main effect of target, F(1, 4) = 0.11, p = 0.76, η_G^2 = 0.0059, nor an interaction between SOA and target, F(9, 36) = 1.00, p = 0.45, η_G^2 = 0.10. Curiously, the precue had no effect on T1 at 400 ms, and this was consistent across observers (**Figure S1**), even though the precue did have an effect on T1 for both shorter and longer SOAs; we have yet to understand this finding.

Third, the overall performance of T1 increased substantially with SOA, from *d'* of ~0.6 at the 100 ms SOA to ~2.1 at the 800 ms SOA (**Figure 1c**). Performance reached levels comparable to 800-ms performance at 250 ms for valid trials and at 400 ms for neutral and invalid trials. We will call this rising function of SOA for T1 "masking-like behavior" [47, 48]. The high T1 performance levels for the longest (800 ms) SOA suggests that memory maintenance was not a limiting factor in the performance of this task.

Fourth, the overall performance of T2 exhibited a dip at intermediate SOAs for all precueing conditions, which reached its lowest point around 300 ms (**Figure 1c**). The dip was larger for invalid trials (reaching *d'* = 1.3 vs. maximum 1.8), but was also present for valid and neutral trials (*d'* = 1.6 vs. maximum 1.9). This U-shaped function of SOA for T2, including its timing, resembles the attentional blink (AB). The AB refers to a difficulty in reporting the second of two targets in a rapid visual sequence, when the targets are 200-500 ms apart [49, 50], and it has been much investigated both experimentally and through modeling [49, 51, 52]. The similarity to the AB includes the so-called "lag-1 sparing," which refers to the fact that T2 performance is not impaired in AB tasks at short SOAs of ~100 ms [53].

Statistically, the variation of *d'* across SOAs and targets was demonstrated by an effect of SOA on *d'*, F(9, 36) = 3.60, p = 0.0028, η_G^2 = 0.19, which differed for the two targets, SOA x target interaction, F(9, 36) = 15.38, p < 0.001, η_G^2 = 0.27.

For RT (**Figure 1e**), there was a trend toward faster T2 responses than T1 responses, F(1, 4) = 5.52, p = 0.078, η_G^2 = 0.091, and the precue influenced RT less for T2 than for T1, validity x target interaction, F(2, 8) = 6.83, p = 0.019, η_G^2 = 0.048. No other main effects or interactions were significant for *d'* or RT, F<1.3.

To summarize, the psychometric time courses for the two-target temporal precueing task were quite rich, with masking-like behavior for T1, AB-like behavior for T2, and the strongest impact of temporal attention on perceptual sensitivity at intermediate SOAs for both targets. These data provide constraints on possible voluntary and involuntary attentional gain dynamics.

Model

General framework. We developed a dynamic perception and attention model in which neural responses are dynamically adjusted through the recurrent processing of a multi-layer neural network. The model describes how perceptual and decision representations evolve over time, through interactions of sensory inputs and attention. The model components are well established in static models of visual cortical function; here we introduced the new dimension of time. Specifically, the model is a generalization of the Reynolds and Heeger (R&H) normalization model of attention [12] into the time domain. We call the present model a "normalization model of dynamic attention."

The model instantiated the hypothesis that the dynamic interactions between attention and orientation perception can be characterized as changes over time in the gain of visual cortical neurons. Gain control is an established mechanism mediating spatial attention [12, 15] and has also been implicated in the effects of rhythmic expectation on perceptual sensitivity [54-56].

Each layer of the model consisted of a population of neurons whose responses followed the R&H equation (**Figure 2**). Each neuron's response was determined by the same basic operations: bottom-up input to a neuron in a given layer was filtered through that neuron's receptive field, multiplied by top-down attentional modulation—which we term *attentional gain*—and then divisively normalized by the activity of its neighbors.

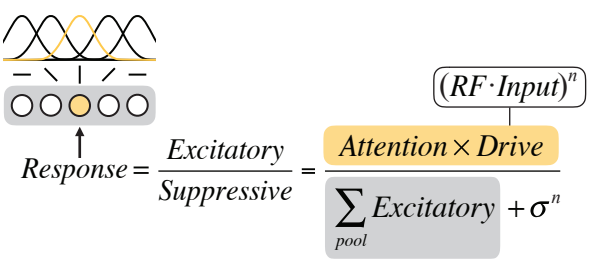

Figure 2. Graphical depiction of the static Reynolds & Heeger normalization model of attention equation. We consider a population of orientation-tuned neurons (illustrated at left, with gray shaded background). The computation of the response of one of the neurons (shaded yellow, with corresponding yellow receptive field (RF)) is shown. The excitatory drive to that neuron is determined by its input drive and attentional modulation (equation text with yellow background). The suppressive drive is determined by the excitatory drives of all the neurons in the local population (equation text with gray background) plus a constant.

To generalize the model from the original static R&H model to a dynamic model, we expressed the model using differential equations that were updated at every time step according to the R&H equation:

$$\tau \frac{d}{dt} r_i = -r_i + \frac{e_i}{s_i + \sigma^n} \quad . \tag{1}$$

Here r_i is the response of a neuron i, where $i \in \{1, \ldots, N\}$ for a population of N neurons; e_i is the excitatory drive to the neuron; s_i is the suppressive drive; σ is a semi-saturation constant that

keeps the denominator from going to zero and controls the neuron's contrast gain; n is a fixed exponent that also contributes to the shape of the contrast response function; and τ is a time constant that determines how long the response takes to rise to steady state when the input turns on and return to zero when it turns off.

The excitatory drive e_i was determined by the equation

$$e_i = a_i \left(\mathbf{w}_i \cdot \mathbf{x} \right)^n , \tag{2}$$

where $\mathbf{x}$ is the bottom-up input to the layer; $\mathbf{w}_i$ is the receptive field (RF) of the neuron; and a_i is top-down attentional gain. Each linear RF computed a weighted sum of its inputs. We describe the inputs $\mathbf{x}$ and RFs $\mathbf{w}$ for each layer in **Methods**.

The suppressive drive s_i was determined by the equation

$$s_i = \sum_{j=1}^{N} e_j , \tag{3}$$

a summation of the excitatory drives of a pool of neurons. Here, the pool was simply all the neurons in the layer (e.g., all orientation preferences at the one spatial location); in general this could be a weighted sum.

At steady state, this differential equation becomes equivalent to the R&H equation; the model therefore retains full generality to predict behavioral and neurophysiological effects of spatial and feature-based attention [19, 20], which have been successfully described and predicted by the R&H model [12].

Model specification. The model architecture was a hierarchical, recurrent neural network, with sensory, attention, and decision layers (**Figure 3a**, **Table S1**). The layers generated continuous neural response (firing rate) time series given attentional precues and continuous stimulus input (**Figure 3b**). Each layer performed the same computation described in "General framework"; only the inputs and outputs were layer-specific. Full details of the model and simulations can be found in **Methods**.

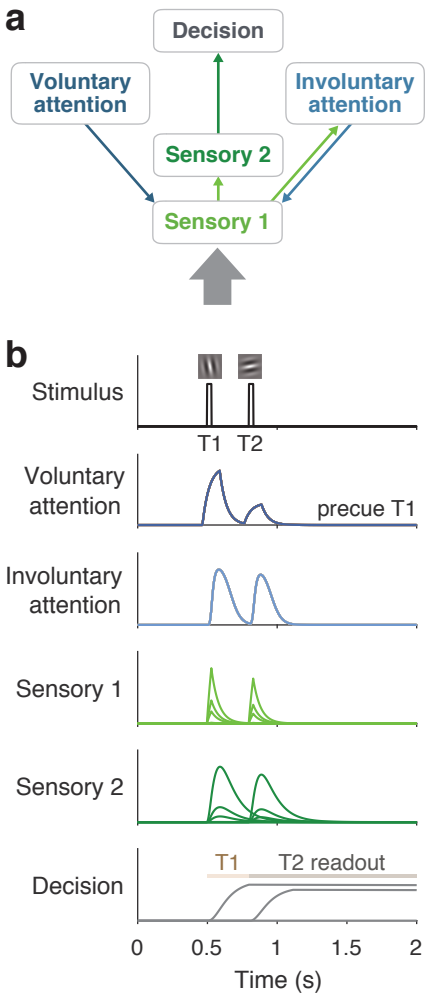

Figure 3. Model architecture and example simulated time series. **a**) Architecture shows sensory input as a thick gray arrow and connections between the layers as thin arrows. An upward arrow indicates input drive, and a downward arrow indicates attentional modulation. **b**) Time series simulated for one trial using a 300 ms SOA, precue to attend to T1, T1 orientation CCW of vertical, and T2 orientation CCW of horizontal, best-fitting parameters. Plots show responses r of each neuron in each layer (separate subplots). Separate lines in the sensory layers are neurons with different orientation preferences, and in the decision layer are neurons reading out T1 and T2 responses. The Decision plot shows decision windows as shaded horizontal lines. Time series are scaled to a uniform amplitude for visualization.

Sensory layers. The sensory layers represented visual cortical areas. Sensory layer 1 (S1) neurons had orientation tuning described by RFs $\mathbf{w}^{S1}$. S1 received stimulus input $\mathbf{x}$: the stimulus orientation (or zero for no stimulus) at every point in time. It also received top-down attentional modulation from both the voluntary and involuntary attention layers. Voluntary and involuntary attention combined multiplicatively to determine the attentional gain a for S1. Sensory layer 2 (S2) received input from S1, inheriting its orientation tuning. S2 had a slower rise and more sustained responses than S1 (because the input to S2 was the output from S1), which helped capture T1 behavioral performance as a function of SOA.

Voluntary attention layer. The voluntary attention layer (VA) increased attentional gain at task-relevant times. Responses in VA depended on the precue (T1, T2, or neutral) and the trial timing. The input to VA was a time-varying control signal (**Figure 4**) that reflected the observer's

knowledge of the precue and SOA. The control signal consisted of square wave pulses around the times of each target. Pulse latency and duration were free parameters. The amplitude of each pulse was determined by the allocation of voluntary attention to each target – i.e., more voluntary attention at a certain time point generated a larger pulse at that time. These control pulses in turn determined the VA response and corresponding attentional gain modulation of S1.

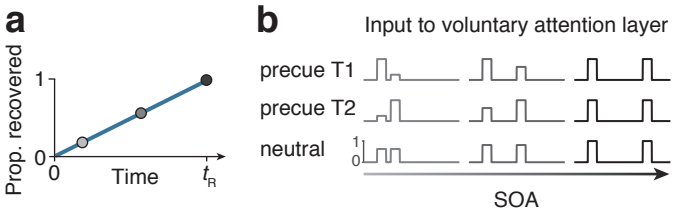

Figure 4. Voluntary attention as a limited but recoverable resource. **a**) Schematic showing the recovery of the input drive to VA, which determined the amplitude of voluntary attentional gain. If maximum voluntary attention was allocated at one moment in time, none was available at the next moment. The available attention then recovered linearly with time, with full recovery at time t_R. **b**) Voluntary attentional control input (y) for different precue conditions and SOAs. When the SOA was short, large attentional tradeoffs between targets occurred due to the limited availability of voluntary attention. When the SOA was long, attention could be allocated maximally to both targets. Voluntary attentional control input was timed, via fitted parameters, to overlap with the sensory responses to each target. In this schematic, w_N = 0.5, so equal attention was allocated to T1 and T2 in the neutral condition.

Voluntary attention was a limited resource across time, generalizing the idea of limited spatial attention resources (e.g., Refs 35, 57, 58) to the time domain (see "Modeling the data"). Immediately after a maximum (=1) allocation of attention, none was available, but over time attention recovered (**Figure 4a**). We modeled the recovery of attention as a linear function of time, with the recovery time given by the parameter t_R. The precue determined the allocation of attention (**Figure 4b**). When the precue was to T1 or T2, maximum attention was allocated to that target, and as much as possible – given the recovery dynamics – was allocated to the other target. When the precue was neutral, a weighting parameter w_N governed the attentional allocation.

Involuntary attention layer. The involuntary attention layer (IA) was stimulus-driven, receiving input from S1. It also fed back to S1, providing a second source of attentional modulation. Because IA responses were driven by S1, they started slightly later than S1 responses (**Figure 3b**). Further, their magnitude depended on the voluntary attentional modulation of S1, because larger S1 responses drove larger IA responses (**Figure 3b**).

Decision layer. The decision layer (D) represented a decision area (e.g., in parietal cortex [59]) and received input from S2. An optimal linear classifier $\mathbf{w}^D$ was used to decode CW vs. CCW evidence at each time step from the S2 population. Because sensory evidence at each time step created a new input drive to D, the layer accumulated evidence across time, similar to drift diffusion models [60]. Decision neurons were target-specific, accumulating evidence only during a corresponding target readout window (**Figure 3b**). The model's task performance was determined by the response of the decision neuron representing the response-cued target at the end of the trial.

Modeling the data

Main model. The normalization model of dynamic attention fit the data well (R^2=0.90) and captured the four main features of the data: (1) voluntary attentional tradeoffs between T1 and

T2, (2) largest precueing effects at intermediate SOAs, (3) masking-like behavior for T1, (4) AB-like behavior for T2 (**Figure 5a**). Fitted parameter values are listed in **Table S2**.

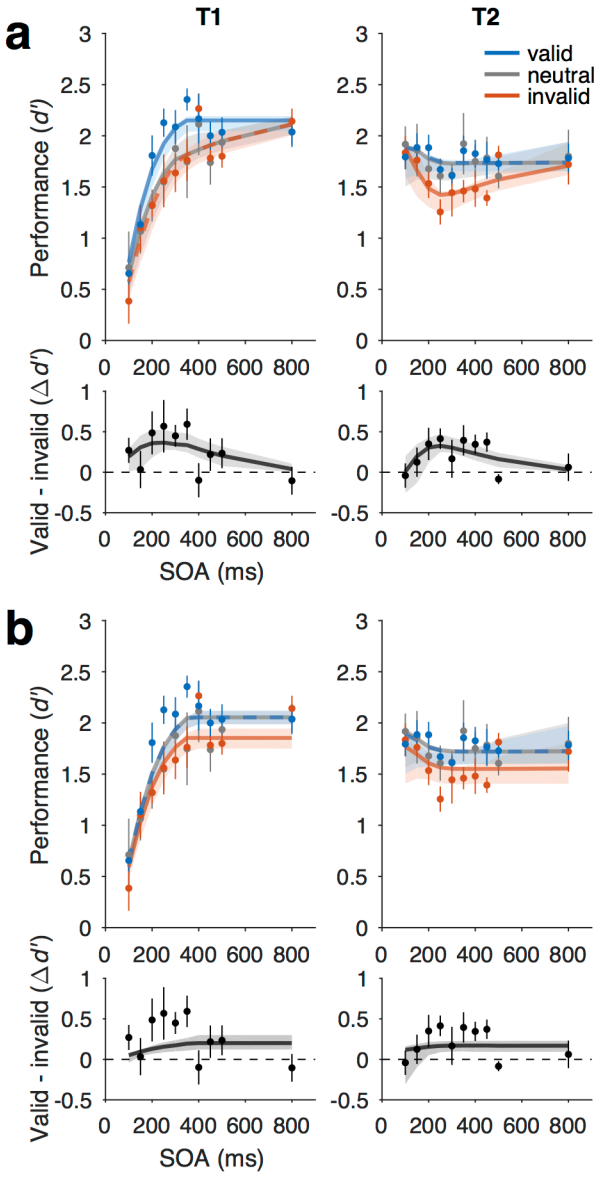

Figure 5. Model fits to perceptual sensitivity data. **a)** Main model with a voluntary attention limit. Top row, performance (*d'*). Bottom row, precueing effect. **b)** No limit model variant. Data points and error bars, behavioral performance (copied from Fig. 4). Curved lines, model fits. Some curves are dashed to reveal overlapping model predictions. Shaded regions, bootstrapped 95% CI for model fits.

To capture the two behavioral features related to voluntary temporal attention – tradeoffs and peak precueing effects at intermediate SOAs – we found it necessary to limit the availability of voluntary attentional gain over time. Specifically, we let voluntary attentional gain be a *limited but recoverable resource* (**Figure 4**). This property generalizes an idea that is standard in the spatial domain to the temporal domain. In the spatial domain, attention to one spatial location leads to improved processing at that location but impaired processing at other locations, relative to a neutral condition [58, 61, 62]. Therefore voluntary spatial attention is considered a limited resource at a single point in time that must be distributed across locations. Analogously, in the

temporal domain, if such a resource is completely used up at one time point, it will not be available at the next time point; but over time, it will recover to its maximum level. Therefore, within the recovery window, the resource must be distributed across sequential items, leading to tradeoffs. Here, the "limited resource" is, concretely, the allocation of voluntary attentional gain. The estimated recovery time of voluntary attention t_R was 918 ms. Additional quantification of the attentional gain dynamics exhibited by the fitted model can be found in **Table S3**.

The overall shapes of the performance functions for T1 and T2 were produced by additional model components. The masking-like behavior for T1 was produced by stopping the decision readout for T1 when T2 appeared. The AB-like behavior for T2 was produced by a combination of three factors: (1) Limited voluntary attention resulted in lower performance at shorter SOAs, especially for invalid trials. (2) At the shortest SOAs (~100 ms), voluntary attention to T1 was sustained long enough to enhance both T1 and T2 sensory responses, boosting T2 performance. (3) Involuntary attentional excitation combined with voluntary attention to further boost T2 performance at the shortest SOAs, resulting in equal, high performance levels across precueing conditions. A model variant without the involuntary attention layer fit the data almost as well (R^2=0.89) and was better in model comparison due to having fewer parameters (ΔAIC=-5.8), although it could not produce AB-like behavior for T2 valid trials (**SI**).

No limit variant. A model without limited voluntary attention (**Figure 5b**) produced a poorer fit (R^2=0.83, ΔAIC=26 with respect to the main model). It also failed to capture the data qualitatively (**Figure 5b**) in two ways. (1) It did not produce tradeoffs in temporal precueing effects: neutral performance was equal to valid performance for both T1 and T2. (2) It did not produce peak precueing effects at intermediate SOAs. Rather, the longest SOAs had maximal precueing effects.

The performance of the no limit model variant reveals why a limit on voluntary attention was necessary. The fact that neutral performance was very similar to valid performance for both targets shows that, without the limit, the model had no incentives to trade off attention between T1 and T2. That is, maximum attention (y=1) could be allocated to both targets on every trial with no performance losses. If more attention to one target had led to worse performance for the other, neutral performance would have been worse than valid performance. Indeed, although we built into this model variant a difference between valid and invalid performance by assuming that the observer would follow the precue to attend to one or both targets, the model would have performed the task better overall if it had ignored the precue and attended to both targets on every trial. In that case, the precue would have had no effect on performance at all, unlike what the data showed.

Other model variants. To further investigate the necessity of limited voluntary attentional gain in this theoretical framework, we developed two alternative model variants designed to produce attentional tradeoff incentives without a limit on voluntary attention (**SI, Figure S2**). One model variant had involuntary attentional inhibition, which suppressed T2 more strongly when T1 was precued. The other variant allowed for mutual normalization of late-stage T1 and T2 responses, such that a stronger T1 response would suppress T2 more. However, when these models were fit to the data, neither of these implementations produced tradeoff profiles. We found that limited voluntary attention was still required to let each of these model variants fit the data (**Figure S3**, **Table S4**), with t_R estimates of 809 ms and 924 ms, respectively.

Discussion

A normalization model of dynamic attention

We developed a normalization model of dynamic attention, generalizing the Reynolds and Heeger normalization model of attention [12] to the time domain. The model is built using components that have support from studies of the visual system and of spatial and feature-based attention, such as linear filters, gain control, rectification, and normalization. Critically, it adds attentional dynamics, i.e., time-varying attentional gain. The model handles temporal attention, including voluntary and involuntary attentional dynamics – in addition to spatial and feature-based attention – in a unified computational framework.

To constrain the model and reveal the dynamics of voluntary temporal attention, we measured how voluntary temporal attention affects perception across time. We found temporal attentional tradeoffs between two sequential targets, which were largest when the targets were separated by SOAs of 200-450 ms. Tradeoffs disappeared at the longest SOAs, revealing a time-limited constraint on processing sequential stimuli that can be accommodated by precisely-timed voluntary control. We also found that the SOA affected the overall performance for the two targets, with masking-like behavior for T1 and AB-like behavior for T2.

The model could reproduce the behavioral data using a combination of voluntary and involuntary attentional gain dynamics, together with a simple implementation of masking. Involuntary attention was estimated to be fast and transient, peaking at 82 ms after stimulus onset, consistent with the dynamics of involuntary spatial attention [33, 58, 63-67]. Although incorporating involuntary attentional gain dynamics into our modeling framework was theoretically motivated, it was not required to explain the current psychophysical data (**SI**). Future research will be needed to further examine whether and how involuntary attention interacts with voluntary temporal attention.

Voluntary attention took the form of a limited resource that recovered over time, with full recovery estimated to take ~1 s. The attentional limitation in our model could be, for example, either a limitation on the available voluntary attentional gain ($\boldsymbol{r}^{VA}$), or a limitation on the activity in voluntary attention control structures (y). The model allowed us to separate the dynamics of voluntary attentional gain from other dynamic processes, such as those related to involuntary attention and those leading to masking-like behavior. It therefore makes specific predictions about different types of attentional gain dynamics (voluntary and involuntary attention time courses in **Figure 3b**). Alternative model variants also required a limitation on voluntary attention across time, but they predicted different gain dynamics (time courses in **Figure S2**). These competing hypotheses could be tested in neurophysiological experiments. The notion of a limited neural resource that can be flexibly allocated is central to multiple domains in psychology and neuroscience, including voluntary spatial attention [61, 62] and working memory [68], but it had not previously been considered for voluntarily attending to specific points in time.

Relation to other attention models concerned with dynamics

Previous modeling frameworks that incorporate both attention and some dynamic element include: the “attention gating model” [69, 70]; the “theory of visual attention” (TVA) [71-73]; and the “competitive interaction theory” [17, 31, 32]. Each framework includes different model variants, some of which incorporate normalization [17, 71]. These models have had success in accounting for behavioral data from various perceptual tasks. Other frameworks focus on rhythmic attention [74, 75], which we do not consider here.

There are several important differences between these models and our dynamic attention model. First, this is the first model of voluntary temporal attention. TVA has been adapted to model a constant level of expectation across time [76, 77], but not attention to specific time points. Second, our model distinguishes between voluntary and involuntary attention, a distinction that is supported by the spatial attention literature [35, 78, 79] and has been reported for temporal attention [74, 80-82]. Third, in previous models [17, 73], the role of attention is to control the encoding of sensory signals into working memory. This view of attention differs from our current model, in which attention modulates sensory signals but has no direct role in encoding. Fourth, our model is built to handle time-varying stimuli and time-varying attention, rather than single, brief displays [31], and without being constrained by attentional episodes [70]. Fifth, our model is explicitly a neural model, built from standard components from visual neuroscience. As such, it makes predictions about the time courses of neural activity that can be tested physiologically.

Application to the attentional blink?

T2 performance in our two-target temporal precueing task resembled T2 performance in AB tasks [49]. The fact that we observed AB-like behavior in a task with no temporal uncertainty, no distractors or masks, and no dual task conditions could help isolate the mechanisms that lead to AB-like behavior [83]. The few AB studies in which voluntary temporal attention has been manipulated have reported inconsistent findings [84-86]. Here, we manipulated voluntary temporal attention and found, in different model variants, that voluntary attentional dynamics could contribute to or be independent from AB-like behavior (**SI**). In our main model, the attentional blink arises from limited voluntary attention. We found no need to invoke other processes previously proposed to contribute to the AB (e.g., working memory limitations, loss of top-down control, alpha oscillations) [51, 87, 88] to explain the AB-like behavior in our task. However, the contribution of such processes is not excluded by our model.

In an influential AB model [52], attention is enhanced by the appearance of a target and suppressed during working memory encoding, which leads to the attentional blink. The initial enhancement of attention is similar to involuntary attentional enhancement in our model but the subsequent suppression differs. A neurophysiological AB model proposes that the AB results from a refractory period in the release of norepinephrine by the locus coeruleus (LC), which limits norepinephrine-driven gain enhancement across time [89]. Future work should examine how voluntary temporal attention affects LC activity; so far there is no evidence that pupil responses, which are influenced by LC, depend on voluntary temporal attention [90].

Future extensions of the model

The current model is a general description of the dynamic interactions between attention and sensory responses. We have focused on how attention affects sensory processing of oriented gratings, a strategy that has proven productive in studies of spatial attention [12, 15, 35, 78, 79, 91], facilitated by our knowledge of how orientation is represented in the visual system [92]. Future work can extend the model to include working memory layers (see **SI**), as sequential processing limitations may also arise in working memory [51], as well as more complex feature representations (using additional sensory layers and different RFs) to handle more complex stimuli. Here the limited resource of voluntary attention was implemented via constraints on the amplitude of the control signal over time and an associated limit on attentional gain. Future work should explore other implementations of a limited resource on voluntary attention over time.

Conclusion

We developed the first model of voluntary and involuntary visual temporal attention, which can serve as a general-purpose computational framework for modeling dynamic attention. Psychophysical measurements revealed perceptual tradeoffs for successive stimuli within sub-second time intervals, which can be controlled by voluntary temporal attention. Precisely timed visual attention may therefore help humans compensate for neural processing limitations over short, behaviorally-relevant timescales. The model predicts that voluntary attentional gain is a limited resource. Future experiments will be needed to test the current model's predictions and specify the attentional gain dynamics related to spatial and feature-based attention. The time-varying nature of the proposed framework – not to mention of vision itself – calls for new data from psychophysical, neurophysiological, and neuroimaging experiments with dynamic displays.

Methods

Behavior

Observers. Five human observers (20-30 years old, 3 female and 2 male) participated in the experiment. All observers provided informed consent, and the University Committee on Activities involving Human Subjects at New York University approved the experimental protocols. Observers were students and researchers at NYU who had experience doing visual psychophysics (though not necessarily temporal attention tasks). All observers had normal or corrected-to-normal vision, and all but author R.N.D. were naïve as to the purpose of the experiment.

Setting. The experiment was conducted in a quiet testing room. During experimental blocks, the only light source was the computer monitor. The experimenter was present to give instructions and throughout training and checked on the observer between testing blocks.

Stimuli. Stimuli were generated on an Apple iMac using MATLAB and Psychophysics Toolbox [93-95] and were displayed on a gamma-corrected Sony Trinitron G520 CRT monitor with a refresh rate of 100 Hz at a viewing distance of 56 cm. Observers' heads were stabilized by a chin-and-head rest. A central white fixation "x" subtended 0.5° visual angle. Visual target stimuli were 4 cpd sinusoidal gratings with a 2D Gaussian spatial envelope (standard deviation 0.7°), presented in the lower right quadrant of the display centered at 5.7° eccentricity (**Figure 1a**). Stimulus contrast was 64%. Placeholders, corners of a 4.25° x 4.25° white square outline (line width 0.08°) centered on the target location, were present throughout the display to minimize spatial uncertainty. The stimuli were presented on a medium gray background (57 cd/m^2). Auditory precues were high (784 Hz; G5) or low (523 Hz; C5) frequency pure tones, or their combination, and were presented via the computer speakers.

Procedure. The task was designed to study the temporal dynamics of voluntary and involuntary temporal attention, including how these two types of attention dynamically interact to affect perception. We used the two-target temporal precueing task developed by Denison et al. [38].

Task. Observers discriminated the orientation of grating patches (**Figure 1**). On each trial, two targets (T1 and T2) were presented at the same spatial location, separated by a fixed stimulus-onset asynchrony (SOA, the time interval between the target onsets) on a given day of testing. The target duration was 30 ms. Each target was tilted slightly clockwise (CW) or counterclockwise (CCW) from either the vertical or horizontal axis, with independent tilts and axes for T1 and T2. Tilts ranged from 1.4° to 2.5° across observers. Both horizontal and vertical

axes were used to discourage observers from adopting a strategy of comparing the two successive targets on a given trial to judge whether they were the same or different.

Overview of experimental manipulations. We manipulated voluntary temporal attention using temporal precues to T1, T2, or both targets. We assumed that the onset of T1 would elicit involuntary temporal attention, and we measured the perceptual effects of involuntary attention on T2 as a function of time by manipulating the SOA. Other time-varying processes that affected the perception of the two targets, like masking, could also be studied as a function of SOA. Because we wanted observers to be able to attend to precise points in time, we eliminated temporal uncertainty by fixing SOA in each testing session, and varied it across sessions. So on a given testing day, the trial timing was constant, and the only thing that varied across trials was the precue to attend to T1, T2, or both targets.

Trial sequence. An auditory precue 1000 ms before T1 instructed observers to attend to one of the targets (informative precue, high tone: attend to T1; low tone: attend to T2) or to attend to both targets (neutral precue, both tones simultaneously). Observers were asked to report the orientation of one of the targets, which was indicated by an auditory response cue 500 ms after T2 (high tone: report T1; low tone: report T2). The duration of the precue and response cue tones was 200 ms. For trials with informative precues (80% of all trials), the response cue matched the precued target with a probability of 75% (valid trials) and the other target with a probability of 25% (invalid trials). For neutral trials (20% of all trials), the two targets were indicated by the response cue with equal probability. To reduce the possibility of speed-accuracy tradeoffs, observers were instructed to withhold their response until the fixation cross dimmed (a “go cue”) 600 ms after the response cue. Observers pressed one of two keys to indicate whether the tilt was CW or CCW relative to the main axis, with unlimited time to respond. Long reaction times (> 2 s) were rare, 0.1% of trials. Reaction times were measured relative to the go cue. Observers received feedback at fixation (correct: green “+”; incorrect: red “-“) after each trial, as well as feedback about performance accuracy (percent correct) following each block of trials.

Sessions. Three observers completed twenty testing sessions (10 SOAs x 2 sessions/SOA, 9.6K trials total), and two observers completed ten sessions (10 SOAs x 1 session/SOA, 4.8K trials total) on separate days. The SOA order was randomly determined for each observer. Observers who completed 2 sessions/SOA did two sets of 10, with a separate random shuffling for each set. Each session consisted of all combinations of precue type (valid: 60%, invalid: 20%, neutral: 20%), probed target (T1, T2), target tilt (CW, CCW; independent for T1 and T2), and target axis (horizontal, vertical; independent for T1 and T2) in a randomly shuffled order, for a total of 480 trials per session. Observers completed 64 practice trials at the start of each session to familiarize them with the SOA for that day.

Training. Observers completed one session of training prior to the experiment to familiarize them with the task and determine their tilt thresholds. Thresholds were found using a 3-down-1-up staircase with all neutral precues at a 250 ms SOA, to achieve an accuracy of ~79% on average across T1 and T2. After determining the tilt threshold, observers completed 64 trials of training with all valid precues, followed by 320 trials identical to an experimental session. The threshold tilt values were used for the remainder of the experiment. (For one observer whose overall performance improved during the first 10 sessions, the tilt was adjusted before the second set of 10 sessions.)

Eye tracking. Eye position was recorded using an EyeLink 1000 eye tracker (SR Research) with a sampling rate of 1000 Hz. Raw gaze positions were converted into degrees of visual angle using the 5-point-grid calibration, which was performed at the start of each experimental run.

Online eye tracking was used to monitor central fixation throughout the experiment. Initiation of each trial was contingent on fixation, with a 750 ms minimum inter-trial interval. Observers were required to maintain fixation, without blinking, from the onset of the precue until the onset of the response cue. If an observer broke fixation during this period, the trial was stopped and repeated at the end of the run.

Statistics. To examine the effects of the experimental manipulations on behavior, we conducted repeated-measures ANOVAs and calculated η_G^2 as a measure of effect size using the *ezANOVA* function from the *ez* package in R. ANOVA assumes normality and sphericity. We tested the normality assumption by fitting a linear model to the data and examining the residuals using Q-Q plots and the Shapiro-Wilk normality test in R. The *d'* measure met the normality assumption. The reaction time measure did not; note, however, that reaction time here was a secondary measure that was analyzed to rule out speed-accuracy tradeoffs, ANOVA is typically robust to violations of normality, and the RT results were consistent with the *d'* results. We confirmed that all significant F tests remained significant after Huynh-Feldt sphericity corrections. All statistical tests were two-sided.

Model

A note on notation: The variables e, s, r, a, g, x, y, and z are all time-varying, e.g., $e(t)$, but for simplicity of notation, we omit the time t. We use boldface, e.g., $\mathbf{x}$, for vectors. As these variables can be both neuron- and layer-specific, we use a subscript to index the neuron and a superscript to refer to the layer, e.g., e_i^{S1} for the excitatory drive to the *i*th neuron of layer S1. When no layer superscript is given, we are referring to the general case, applicable to all layers. For constants, which are neither time-varying nor neuron-specific, we use only subscripts.

Model specification. The model consisted of a hierarchical, recurrent neural network, with different layers (**Figure 3a**): sensory layers, analogous to visual cortical areas; attention layers, which modulated the sensory responses; and a decision layer, which read out the sensory responses and reported a task decision (here, clockwise vs. counterclockwise grating tilt). All model parameters are listed in **Table 1**.

Parameter	Description	Value	95% CI
All layers			
n	exponent	1.5	--
Sensory layer 1			
τ_{S1}	time constant	52	[49 116]
σ_{S1}	semi-saturation constant	1.4	[1.2 2.0]
Sensory layer 2			
τ_{S2}	time constant	100	[69 120]
σ_{S2}	semi-saturation constant	0.1	--
Decision layer			
τ_D	time constant	1e5	--
σ_D	semi-saturation constant	0.7	--

Voluntary attention layer			
τ_{VA}	time constant	50	--
σ_{A} (shared with IA)	semi-saturation constant	20	--
b_{VA}	amplitude of voluntary gain modulation	40	[21 50]
t_{VAOn}	latency of voluntary control signal onset	-34	[-223 -6]
t_{VADur}	duration of voluntary control signal	124	[99 374]
t_{R}	recovery time of voluntary gain	918	[600 1091]
w_{N}	weight to treat neutral precue more like precue T1 (1) or precue T2 (0)	0.28	[0.01 0.53]
Involuntary attention layer			
τ_{IA}	time constant	2	--
σ_{A} (shared with VA)	semi-saturation constant	20	--
b_{IA}	amplitude of involuntary gain modulation	8.5	[0.8 27.9]
$h_{\mathrm{IA}}{:}p$	shape parameter for involuntary temporal prefilter	2.2	[0.04 49.9]
$h_{\mathrm{IA}}{:}q$	scaling parameter for involuntary temporal prefilter	0.023	[0.01 0.09]
Fitting			
s_{T1}	scaling constant to relate model output to *d'* for T1	1	--
s_{T2}	scaling constant to relate model output to *d'* for T2	0.80	[0.77 0.84]
Number of parameters			
Total		20	--
Fitted		12	--

Table 1. Model parameters. Light gray shading indicates that the parameter was fixed to a set value and not optimized during fitting. All times (i.e., time constants, latencies) are given in ms. Negative latencies for t_{VAOn} mean that the voluntary control signal started before the target.

Sensory layer 1. The first sensory layer (**Figure 3b**, S1) represented an early-stage visual area and received stimulus input at every time step. The excitatory drive for each neuron i was

$$e_i^{\mathrm{S1}} = a_i^{\mathrm{S1}} \left(\mathbf{w}_i^{\mathrm{S1}} \cdot \mathbf{x} \right)^n . \tag{4}$$

The stimulus $\mathbf{x}$ was represented in terms of its orientation: a vector of length M, with each element corresponding to a different orientation θ. When the target grating was on, the element corresponding to the stimulus orientation had value c, the stimulus contrast; $\mathbf{x} = (0, 0, ..., c, ..., 0)$. When the grating was off, all elements were zero-valued. We showed the model the same 2-target trial sequences (**Figure 3b**, Stimulus) as we showed observers (**Figure 1b**).

Layer S1 had 12 orientation-selective model neurons (each of which could represent a larger neural population), which tiled orientation at a single spatial location. Each RF $\mathbf{w}_i^{S1} = (w_{i,1}, w_{i,2}, ..., w_{i,M})$, was designed to be an orientation tuning curve described by one cycle of a raised cosine,

$$w_{i,j} = \left| \cos\left(\theta_j - \varphi_i\right)^m \right|, \tag{5}$$

where θ is orientation, with sampled orientations indexed by j, and φ_i is the preferred orientation of the ith neuron. Preferred orientations were evenly spaced, φ = $(0, \pi/N, ..., \pi\text{-}\pi/N)$. The exponent m, which governs the width of the tuning curves, was set to m = $2N$-1. These tuning curves ensure even tiling of orientation space. Orientation selectivity was all that was needed to model our task, but in general the model neurons would be selective also for spatial location and spatial frequency.

S1 received top-down attentional gain modulation from both the voluntary (VA) and involuntary (IA) attention layers, whose responses independently and multiplicatively modulated the sensory drive. The attentional gain a^{S1} was

$$a_i^{S1} = \left\lfloor 1 + b_{VA} r_i^{VA} \right\rfloor \left\lfloor 1 + b_{IA} r_i^{IA} \right\rfloor, \tag{6}$$

where $\mathbf{r}^{VA}$ and $\mathbf{r}^{IA}$ are the responses of the voluntary and involuntary attention layers, respectively; b_{VA} and b_{IA} are free parameters determining the amplitude of voluntary and involuntary attentional modulation; and brackets denote halfwave rectification, which ensures that attentional modulation is positive. The baseline attentional gain was assumed to be 1, which left sensory responses unchanged in the absence of top-down attentional modulation. When a was greater than one, sensory responses increased above baseline; we call this "excitatory" attentional modulation. When a was less than one, sensory responses decreased below baseline; we call this "inhibitory" attentional modulation.

Sensory layer 2. The second sensory layer (**Figure 3b**, S2) represented a later-stage visual area and received input from S1. Layer S2 also had 12 neurons, and each neuron received input from a single S1 neuron, thereby inheriting the orientation tuning of the S1 neuron ($\mathbf{x}$ = $\mathbf{r}^{S1}$). S2 did not receive attentional modulation, so its attentional gain a was effectively fixed to 1. The excitatory drive for each neuron i was therefore simply

$$e_i^{S2} = \left(r_i^{S1}\right)^n, \tag{7}$$

where corresponding neurons i in S2 and S1 have the same orientation preference. S2 responses are also determined by Equation 1, which includes a temporal low-pass filter with time constant τ_{S2}. Therefore, S2 had a slower rise and more sustained responses than S1, which helped capture T1 behavioral performance as a function of SOA. However, temporal RFs with more complex dynamics within a layer (e.g., cascade of exponentials) could have achieved the same result using only one sensory layer.

Decision layer. The decision layer (**Figure 3b**, Decision) represented a decision area (e.g., in parietal cortex [59]) and received input from S2. To encode information about temporal order, there were two neurons in the decision layer, one for T1 and one for T2. A binary (0 or 1) decision window g gated the input to each neuron: a neuron received input drive only when its decision window was on. The T1 neuron's decision window started at the onset of T1 and stopped at the onset of T2. The T2 neuron's decision window started at the onset of T2 and stopped at the end of the trial. Thus, the decision layer read out from successive decision windows for the two targets (**Figure 3b**, Decision, shaded regions). The decision window cutoff

for T1 implemented a simplified version of masking, standing in for a mechanism that curtails T1-related signals when T2 appears [47, 96].

The input drive to the decision layer was the evidence for clockwise (CW) or counterclockwise (CCW) tilts, based on the responses of S2. The excitatory drive for each neuron i was

$$e_i^{\mathrm{D}} = g_i\left(\mathbf{w}_i^{\mathrm{D}} \cdot \mathbf{r}^{\mathrm{S2}}\right). \tag{8}$$

The inner product in this equation represents an optimal linear classifier that decoded CW vs. CCW evidence at each time step from the S2 population response. The classifier projected $\mathbf{r}^{\mathrm{S2}}$ onto the difference $\mathbf{w}^{\mathrm{D}}$ between two templates corresponding to ideal responses to the two possible stimuli, CW and CCW. These templates were "ideal" in the sense that they were the population responses for CW or CCW stimuli at full contrast and with no noise. (We assumed that the orientation axis –vertical or horizontal– was known to the observer, so the comparison was only between the CW and CCW templates on the relevant orientation axis.) The classifier projection gave a continuous value that indicated the similarity of the population response on that trial to the CW vs. CCW templates. We arbitrarily assigned CW evidence to positive values and CCW evidence to negative values; an alternative implementation could have used different neurons to represent evidence for each choice. The sign of the evidence indicated a CW vs. CCW choice, and the magnitude indicated the strength of evidence for the choice. Because sensory evidence at each time step created a new input drive to the decision layer, the decision layer accumulated evidence across time, similar to drift diffusion models [60]. We fixed the time constant of the decision layer to a large value to minimize integration leak and allow faithful evidence accumulation across the trial.

Depending on the response cue, the T1 or T2 decision neuron's response at the end of the trial was used to determine the model's performance. Specifically, the model's CW vs. CCW choice on each trial was determined by the sign of the response, and the magnitude of the decision neuron's response was presumed to be proportional to the experimentally observed *d'* using a fixed scaling factor. This corresponds to a maximum-likelihood decision assuming additive Gaussian noise. We also fit a separate, relative scaling factor for T2, as the maximum *d'* was lower for T2 than T1. We do not have an explanation for overall differences between T1 and T2, which we have found to vary across datasets and individuals [38, 39], but we let these differences be captured by the T2 scaling parameter.

Voluntary attention layer. Responses in VA (**Figure 3b**) depended on the precue for a given trial (e.g., precue T1) and knowledge of task timing, so that VA responses increased at the time of the relevant sensory responses. The input y to the layer was a time-varying control signal that reflected the observer's knowledge of the SOA and the precue type ∈ {T1, T2, neutral}. The excitatory drive was determined by the input,

$$e_i^{\mathrm{VA}} = y^n. \tag{9}$$

The control signal was a square wave at each target time (**Figure 4**). The timing of each square wave was determined by the SOA and two free parameters, t_{VAOn} and t_{VADur}, which controlled the onset and duration of the square wave. The amplitude of each square wave was determined by the allocation of voluntary attentional gain to each target.

In one model variant, voluntary attentional gain was a limited resource across time. The amplitudes of the square waves then depended on two additional free parameters: a time

constant t_R and an attentional weighting parameter w_N. We implemented the limited resources idea by assuming that immediately after a maximum allocation of gain, no gain would be available, but over time the available gain would recover to the maximum level (**Figure 4**). We modeled the recovery of attention as a linear function of time, with the recovery time given by the parameter t_R. We defined the maximum attention allocation (the amplitude of y) at a given time to be 1. Thus the total attention available to both targets for a given SOA was

$$\text{total attention}_{\text{SOA}} = 1 + \min\left(\frac{\text{SOA}}{t_{\text{R}}}, 1\right). \tag{10}$$

The total attention available across the SOA was distributed across the two targets according to the precue. When the precue was informative (directing attention to one target), the maximum attention for a single target (=1) was allocated to the precued target, and the remainder from the total available attention across both targets was allocated to the other target. (Similar results were obtained when a weighting parameter determined the proportion of attention allocated to the precued target.) For example, if the t_R was 1000 ms, and the SOA was 400 ms, the total attention available across both targets was 1 (to the precued target) + 400/1000 (to the other target) = 1.4. When the precue was neutral, a weighting parameter w_N governed the attentional allocation. Due to a perceived or actual asymmetry between the two targets, observers may have a tendency to treat a neutral precue more like a precue for T1 or for T2; w_N captured this possible bias. Again, attention in excess of 1 to a given target was shifted to the other target to make full use of the available attention. **Figure 4** shows examples of the attention control input to VA.

Involuntary attention layer. Responses in IA were driven by input from S1 (i.e., "stimulus driven"), as shown in **Figure 3**. The excitatory drive was

$$e_i^{\text{IA}} = \left(\mathbf{w}^{\text{IA}} \cdot \mathbf{z}\right)^n, \tag{11}$$

where $\boldsymbol{z}$ is a temporally filtered version of $\boldsymbol{r}^{S1}$. The temporal filter, which we refer to as a "prefilter", was a gamma function (h_{IA}) with amplitude fixed to 1 and fitted shape and scaling parameters p and q. The gamma function is equivalent to a cascade of exponential (low-pass) filters. The RF $\mathbf{w}^{IA}$ was the same for all the neurons in the layer and weighted all S1 responses equally, i.e., it was not feature-selective. Because IA responses are driven by S1, they start slightly later than S1 responses (**Figure 3b**). Further, their magnitude depends on the voluntary attentional modulation of S1, with stronger involuntary responses to sensory responses that are voluntarily attended (**Figure 3b**).

Model variants. In addition to the main model just described, several variants were developed and tested.

No limit model variant. To test whether the limit on voluntary attentional gain was needed to explain the behavioral data, we created a "no limit" alternative. The only difference between the original model variant and the no limit variant was that, in the no limit variant, voluntary attentional gain could be allocated at the maximum level (still defined as 1) at any time, regardless of its allocation at other times. Therefore, this model variant did not require the parameters that determined the allocation of limited voluntary attentional gain in the main model variant: t_R, and w_N.

To give the no limit model variant the best chance to capture the data, we assumed that the model was an "obedient" observer that followed the instruction of the precue to attend to T1, attend to T2, or attend to both targets (neutral precue). So for precue T1 trials, the model allocated voluntary gain maximally (y=1) to T1 and minimally (y=0) to T2, and vice versa for precue T2 trials. For neutral trials, the model allocated voluntary gain maximally (y=1) to both targets. Note that if we had allowed the model to adopt an optimal strategy – that is, to maximize performance accuracy – it would have attended maximally to all targets even for invalid trials, resulting in no difference between valid and invalid trials. This strategy was not available to the variant with limits on voluntary attention.

Other model variants. We tested additional variants of the model to assess whether other mechanisms could explain the behavioral data without a temporal limit on voluntary attention. These model variants are described in **SI**.

Simulation procedures. The model simulations were run using MATLAB. Each simulated trial lasted 2.1 s with time steps Δt of 2 ms. The continuous differential equation representing the dynamical version of the R&H normalization equation (**Equation 1**) was discretized as

$$\mathbf{r}(t) = \mathbf{r}(t-\Delta t) + \frac{\Delta t}{\tau}\left(-\mathbf{r}(t-1) + \frac{\mathbf{e}(t)}{\mathbf{s}(t)+\sigma^n}\right). \tag{12}$$

To generate the model performance for one condition (defined by a combination of precue and SOA), a single trial was simulated recursively for 1050 time steps t. The only time series specified in advance were the stimulus time series, the voluntary attention control time series y, and the decision window time series g specifying when evidence would be accumulated for each target. Voluntary attention control was specified based on the precue. The other time-varying quantities – the excitatory drives e, suppressive drives s, and neuronal responses r for each layer – were calculated at each time step, based on the values of the other time-varying quantities at the current and, when prefiltering was applied, previous time steps.

To generate the full psychometric functions containing 60 data points (3 precues x 10 SOAs x 2 targets), each condition was simulated once with no noise to obtain the model performance for both targets in that condition. We performed all simulations with the T1 stimulus CCW from vertical and the T2 stimulus CCW from horizontal, which produced the same behavior as the average across all possible stimulus sequences.

Fitting procedures. We fit each model variant to the group average *d'* data (60 data points). Model fitting was conducted in two phases. In the first phase, the cost function – sum of squared error between model output and data – was evaluated at 2,000 parameter sets sampled from reasonable parameter ranges. To sample evenly across the full range, each range was divided into 400 bins, and 5 parameter values were sampled uniformly from each bin. In the second phase, the 40 parameter sets with the lowest cost from the first phase were used as starting points for optimization. The optimization algorithm was Bayesian adaptive direct search (BADS) [97], which is well-suited for the number of parameters and cost function evaluation time of our model. The optimization producing the lowest cost across all starting points was selected as the best fit.

Some parameters were fixed by hand (i.e., not fit) for theoretical reasons or to minimize redundancies among the fitted parameters. Fixed parameters are shaded in gray in **Table 1**. The values of all fixed and best-fit parameters are given in **Table 1**.

Resampling procedures. We obtained confidence intervals on the parameter estimates and model predictions by bootstrapping the data and refitting the model 100 times. For each bootstrap, we aggregated all trials from all observers for each condition, resampled the trials with replacement, and recalculated *d'*. Then, for each resampled dataset, we performed the fitting procedure as described above (2,000 initial cost function evaluations followed by optimization from 40 starting points). This procedure yielded 100 fits of resampled data. Confidence intervals on parameter estimates (**Table 1**) were calculated from the bootstrapped estimates. Confidence intervals on model fits (**Figure 5**) were calculated for each SOA using the bootstrapped model output (predicted *d'*).

Parameter interpretation. Several parameters contributed to the attentional dynamics exhibited by each model variant, including the attentional time constants, onset and duration of the voluntary attention control signal, temporal prefilter for the involuntary attention layer, and amplitudes of voluntary and involuntary attentional modulation. We used these fitted parameters to calculate summary metrics describing the voluntary and involuntary attentional dynamics produced by the model. Specifically, we calculated the peak gain amplitudes, the latencies of those peak amplitudes, and the maximum durations of the gain modulations. To calculate the duration of an attentional gain response, we defined a response as non-zero whenever its absolute value was greater than 1% of the maximum response. Note that gain amplitudes are in arbitrary units that should only be compared to the other amplitudes from a given fit.

References

1. Carandini, M. & Heeger, D.J. Normalization as a canonical neural computation. *Nature Reviews Neuroscience* **13**, 51-62 (2012).
2. Heeger, D.J. Normalization of cell responses in cat striate cortex. *Visual Neuroscience* **9**, 181-197 (1992).
3. Bonin, V., Mante, V. & Carandini, M. The suppressive field of neurons in lateral geniculate nucleus. *Journal of Neuroscience* **25**, 10844-10856 (2005).
4. Busse, L., Wade, A.R. & Carandini, M. Representation of concurrent stimuli by population activity in visual cortex. *Neuron* **64**, 931-942 (2009).
5. Carandini, M. & Heeger, D.J. Summation and division by neurons in primate visual cortex. *Science* **264**, 1333-1336 (1994).
6. Ni, A.M. & Maunsell, J.H.R. Spatially tuned normalization explains attention modulation variance within neurons. *Journal of Neurophysiology* **118**, 1903-1913 (2017).
7. Li, H.-H., Carrasco, M. & Heeger, D.J. Deconstructing interocular suppression: attention and divisive normalization. *PLoS Computational Biology* **11**, e1004510 (2015).
8. Li, H.-H., Rankin, J., Rinzel, J., Carrasco, M. & Heeger, D.J. Attention model of binocular rivalry. *Proceedings of the National Academy of Sciences* **114**, E6192-E6201 (2017).
9. Ling, S. & Blake, R. Normalization regulates competition for visual awareness. *Neuron* **75**, 531-540 (2012).
10. Louie, K., LoFaro, T., Webb, R. & Glimcher, P.W. Dynamic divisive normalization predicts time-varying value coding in decision-related circuits. *Journal of Neuroscience* **34**, 16046-16057 (2014).
11. Ohshiro, T., Angelaki, D.E. & Deangelis, G.C. A normalization model of multisensory integration. *Nature Neuroscience* **14**, 775-782 (2011).
12. Reynolds, J.H. & Heeger, D.J. The normalization model of attention. *Neuron* **61**, 168-185 (2009).
13. Boynton, G.M. A framework for describing the effects of attention on visual responses. *Vision Research* **49**, 1129-1143 (2009).
14. Lee, J. & Maunsell, J.H.R. A normalization model of attentional modulation of single unit responses. *PLoS ONE* **4**, e4651 (2009).
15. Maunsell, J.H.R. Neuronal mechanisms of visual attention. *Annual Review of Vision Science* **1**, 373-391 (2015).
16. Schwedhelm, P., Krishna, B.S. & Treue, S. An extended normalization model of attention accounts for feature-based attentional enhancement of both response and coherence gain. *PLoS Computational Biology* **12**, e1005225 (2016).
17. Smith, P.L., Sewell, D.K. & Lilburn, S.D. From shunting inhibition to dynamic normalization: attentional selection and decision-making in brief visual displays. *Vision Research* **116**, 219-240 (2015).
18. Ni, A.M. & Maunsell, J.H.R. Neuronal effects of spatial and feature attention differ due to normalization. *Journal of Neuroscience* **39**, 5493-5505 (2019).
19. Herrmann, K., Heeger, D.J. & Carrasco, M. Feature-based attention enhances performance by increasing response gain. *Vision Research* **74**, 10-20 (2012).
20. Herrmann, K., Montaser-Kouhsari, L., Carrasco, M. & Heeger, D.J. When size matters: attention affects performance by contrast or response gain. *Nature Neuroscience* **13**, 1554-1559 (2010).
21. Zhang, X., Japee, S., Safiullah, Z., Mlynaryk, N. & Ungerleider, L.G. A normalization framework for emotional attention. *PLoS Biology* **14**, e1002578 (2016).
22. Carandini, M., Heeger, D.J. & Movshon, J.A. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience* **17**, 8621-8644 (1997).

23. Reynaud, A., Masson, G.S. & Chavane, F. Dynamics of local input normalization result from balanced short- and long-range intracortical interactions in area V1. *Journal of Neuroscience* **32**, 12558-12569 (2012).
24. Sit, Y.F., Chen, Y., Geisler, W.S., Miikkulainen, R. & Seidemann, E. Complex dynamics of V1 population responses explained by a simple gain-control model. *Neuron* **64**, 943-956 (2009).
25. Zhou, J., Benson, N.C., Kay, K. & Winawer, J. Compressive temporal summation in human visual cortex. *Journal of Neuroscience*, 1724-1717 (2017).
26. Wainwright, M.J., Schwartz, O. & Simoncelli, E.P. Natural image statistics and divisive normalization: modeling nonlinearities and adaptation in cortical neurons. in *Statistical Theories of the Brain* (ed. R.P. Rao, B. Olshausen & M. Lewicki) 1-22 (MIT Press, 2002).
27. Westrick, Z.M., Heeger, D.J. & Landy, M.S. Pattern adaptation and normalization reweighting. *Journal of Neuroscience* **36**, 9805-9816 (2016).
28. Wilson, H.R. & Humanski, R. Spatial frequency adaptation and contrast gain control. *Vision Research* **33**, 1133-1149 (1993).
29. Wissig, S.C. & Kohn, A. The influence of surround suppression on adaptation effects in primary visual cortex. *Journal of Neurophysiology* **107**, 3370-3384 (2012).
30. Kaliukhovich, D.A. & Vogels, R. Divisive normalization predicts adaptation-induced response changes in macaque inferior temporal cortex. *Journal of Neuroscience* **36**, 6116-6128 (2016).
31. Smith, P.L. & Sewell, D.K. A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychological Review* **120**, 589-627 (2013).
32. Smith, P.L. & Ratcliff, R. An integrated theory of attention and decision making in visual signal detection. *Psychological Review* **116**, 283-317 (2009).
33. Müller, H.J. & Rabbitt, P.M. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance* **15**, 315-330 (1989).
34. Carrasco, M., Ling, S. & Read, S. Attention alters appearance. *Nature Neuroscience* **7**, 308-313 (2004).
35. Carrasco, M. Visual attention: The past 25 years. *Vision Research* **51**, 1484-1525 (2011).
36. Nobre, A.C. & van Ede, F. Anticipated moments: temporal structure in attention. *Nature Reviews Neuroscience* **19**, 34-48 (2018).
37. Correa, A., Lupiáñez, J. & Tudela, P. Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychonomic Bulletin & Review* **12**, 328-334 (2005).
38. Denison, R.N., Heeger, D.J. & Carrasco, M. Attention flexibly trades off across points in time. *Psychonomic Bulletin & Review* **24**, 1142-1151 (2017).
39. Fernández, A., Denison, R.N. & Carrasco, M. Temporal attention improves perception similarly at foveal and parafoveal locations. *Journal of Vision* **19**, 12 (2019).
40. Rohenkohl, G., Gould, I.C., Pessoa, J. & Nobre, A.C. Combining spatial and temporal expectations to improve visual perception. *Journal of Vision* **14** (2014).
41. Samaha, J., Bauer, P., Cimaroli, S. & Postle, B.R. Top-down control of the phase of alpha-band oscillations as a mechanism for temporal prediction. *Proceedings of the National Academy of Sciences* **112**, 8439-8444 (2015).
42. Anderson, B. & Sheinberg, D.L. Effects of temporal context and temporal expectancy on neural activity in inferior temporal cortex. *Neuropsychologia* **46**, 947-957 (2008).
43. Correa, A., Lupiáñez, J., Madrid, E. & Tudela, P. Temporal attention enhances early visual processing: a review and new evidence from event-related potentials. *Brain Research* **1076**, 116-128 (2006).

44. Coull, J.T. & Nobre, A.C. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience* **18**, 7426-7435 (1998).
45. Miniussi, C., Wilding, E.L., Coull, J.T. & Nobre, A.C. Orienting attention in time. Modulation of brain potentials. *Brain* **122 (Pt 8)**, 1507-1518 (1999).
46. Denison, R.N., Yuval-Greenberg, S. & Carrasco, M. Directing voluntary temporal attention increases fixational stability. *Journal of Neuroscience* **39**, 353-363 (2019).
47. Breitmeyer, B. & Ogmen, H. *Visual Masking* (Oxford University Press, Oxford, 2006).
48. Kahneman, D. Method, findings, and theory in studies of visual masking. *Psychological Bulletin* **70**, 404-425 (1968).
49. Dux, P.E. & Marois, R. The attentional blink: a review of data and theory. *Attention, Perception, & Psychophysics* **71**, 1683-1700 (2009).
50. Raymond, J.E., Shapiro, K.L. & Arnell, K.M. Temporary suppression of visual processing in an RSVP task: an attentional blink? . *Journal of Experimental Psychology: Human Perception and Performance* **18**, 849-860 (1992).
51. Chun, M.M. & Potter, M.C. A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance* **21**, 109-127 (1995).
52. Wyble, B., Potter, M.C., Bowman, H. & Nieuwenstein, M. Attentional episodes in visual perception. *Journal of Experimental Psychology: General* **140**, 488-505 (2011).
53. Potter, M.C., Chun, M.M., Banks, B.S. & Muckenhoupt, M. Two attentional deficits in serial target search: the visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning Memory and Cognition* **24**, 979-992 (1998).
54. Auksztulewicz, R., Myers, N.E., Schnupp, J.W. & Nobre, A.C. Rhythmic Temporal Expectation Boosts Neural Activity by Increasing Neural Gain. *Journal of Neuroscience* **39**, 9806-9817 (2019).
55. Cravo, A.M., Rohenkohl, G., Wyart, V. & Nobre, A.C. Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *Journal of Neuroscience* **33**, 4002-4010 (2013).
56. Rohenkohl, G., Cravo, A.M., Wyart, V. & Nobre, A.C. Temporal expectation improves the quality of sensory information. *Journal of Neuroscience* **32**, 8424-8428 (2012).
57. Luck, S.J., Hillyard, S.A., Mouloua, M. & Hawkins, H.L. Mechanisms of visual-spatial attention: resource allocation or uncertainty reduction? *Journal of Experimental Psychology: Human Perception and Performance* **22**, 725-737 (1996).
58. Pestilli, F. & Carrasco, M. Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vision Research* **45**, 1867-1875 (2005).
59. Gold, J.I. & Shadlen, M.N. The neural basis of decision making. *Annual Review of Neuroscience* **30**, 535-574 (2007).
60. Ratcliff, R., Smith, P.L., Brown, S.D. & McKoon, G. Diffusion decision model: current issues and history. *Trends in Cognitive Sciences* **20**, 260-281 (2016).
61. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* **18**, 193-222 (1995).
62. Giordano, A.M., McElree, B. & Carrasco, M. On the automaticity and flexibility of covert attention: a speed-accuracy trade-off analysis. *Journal of Vision* **9**, 30 31-10 (2009).
63. Cheal, M., Lyon, D.R. & Hubbard, D.C. Does attention have different effects on line orientation and line arrangement discrimination? *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology* **43**, 825-857 (1991).
64. Hein, E., Rolke, B. & Ulrich, R. Visual attention and temporal discrimination: differential effects of automatic and voluntary cueing. *Visual Cognition* **13**, 29-50 (2006).
65. Ling, S. & Carrasco, M. Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision Research* **46**, 1210-1220 (2006).

66. Nakayama, K. & Mackeben, M. Sustained and transient components of focal visual attention. *Vision Research* **29**, 1631-1647 (1989).
67. Remington, R.W., Johnston, J.C. & Yantis, S. Involuntary attentional capture by abrupt onsets. *Perception and Psychophysics* **51**, 279-290 (1992).
68. Ma, W.J., Husain, M. & Bays, P.M. Changing concepts of working memory. *Nature Neuroscience* **17**, 347-356 (2014).
69. Reeves, A. & Sperling, G. Attention gating in short-term visual memory. *Psychological Review* **93**, 180-206 (1986).
70. Sperling, G. & Weichselgartner, E. Episodic theory of the dynamics of spatial attention. *Psychological Review* **102**, 503 (1995).
71. Bundesen, C. A theory of visual attention. *Psychological Review* **97**, 523-547 (1990).
72. Bundesen, C., Habekost, T. & Kyllingsbæk, S. A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological Review* **112**, 291-328 (2005).
73. Bundesen, C., Vangkilde, S. & Petersen, A. Recent developments in a computational theory of visual attention (TVA). *Vision Research* **116**, 210-218 (2015).
74. Jones, M.R. *Time Will Tell: A Theory of Dynamic Attending* (Oxford University Press, New York, NY, 2019).
75. Large, E.W. & Jones, M.R. The dynamics of attending: How people track time-varying events. *Psychological Review* **106**, 119-159 (1999).
76. Vangkilde, S., Coull, J.T. & Bundesen, C. Great expectations: temporal expectation modulates perceptual processing speed. *Journal of Experimental Psychology: Human Perception and Performance* **38**, 1183-1191 (2012).
77. Vangkilde, S., Petersen, A. & Bundesen, C. Temporal expectancy in the context of a theory of visual attention. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, 20130054 (2013).
78. Anton-Erxleben, K. & Carrasco, M. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nature Reviews Neuroscience* **14**, 188-200 (2013).
79. Carrasco, M. & Barbot, A. How attention affects spatial resolution. *Cold Spring Harbor Symposia on Quantitative Biology* **79**, 149-160 (2015).
80. Lawrence, M.A. & Klein, R.M. Isolating exogenous and endogenous modes of temporal attention. *Journal of Experimental Psychology: General* **142**, 560-572 (2013).
81. McCormick, C.R., Redden, R.S., Lawrence, M.A. & Klein, R.M. The independence of endogenous and exogenous temporal attention. *Attention, Perception, & Psychophysics* **80**, 1885-1891 (2018).
82. Moon, J., Choe, S., Lee, S. & Kwon, O.S. Temporal Dynamics of Visual Attention Allocation. *Scientific reports* **9**, 3664 (2019).
83. Nieuwenstein, M., Van der Burg, E., Theeuwes, J., Wyble, B. & Potter, M. Temporal constraints on conscious vision: on the ubiquitous nature of the attentional blink. *Journal of Vision* **9**, 18.11-14 (2009).
84. Hilkenmeier, F. & Scharlau, I. Rapid allocation of temporal attention in the attentional blink paradigm. *European Journal of Cognitive Psychology* **22**, 1222-1234 (2010).
85. Martens, S. & Johnson, A. Timing attention: cuing target onset interval attenuates the attentional blink. *Memory & Cognition* **33**, 234-240 (2005).
86. Visser, T.A.W., Tang, M.F., Badcock, D.R. & Enns, J.T. Temporal cues and the attentional blink: A further examination of the role of expectancy in sequential object perception. *Attention, Perception, & Psychophysics* **76**, 2212-2220 (2014).
87. Di Lollo, V., Kawahara, J.-i., Shahab Ghorashi, S.M. & Enns, J.T. The attentional blink: resource depletion or temporary loss of control? *Psychological Research* **69**, 191-200 (2005).
88. Shapiro, K.L., Hanslmayr, S., Enns, J.T. & Lleras, A. Alpha, beta: The rhythm of the attentional blink. *Psychonomic Bulletin & Review* **34**, 1472-1478 (2017).

89. Nieuwenhuis, S., Gilzenrat, M.S., Holmes, B.D. & Cohen, J.D. The role of the locus coeruleus in mediating the attentional blink: a neurocomputational theory. *Journal of Experimental Psychology: General* **134**, 291-307 (2005).
90. Denison, R.N., Parker, J.A. & Carrasco, M. Modeling pupil responses to rapid sequential events. *bioRxiv* **655902** (2019).
91. Carrasco, M. Spatial attention: Perceptual modulation. in *The Oxford Handbook of Attention* (ed. S. Kastner & A.C. Nobre) 183-230 (Oxford University Press, 2014).
92. DeValois, R.L. & DeValois, K.K. *Spatial Vision* (Oxford University Press, 1990).
93. Brainard, D.H. The Psychophysics Toolbox. *Spatial Vision* **10**, 433-436 (1997).
94. Kleiner, M., Brainard, D.H. & Pelli, D.G. What's new in Psychtoolbox-3? ECVP Abstract Supplement *Perception* **36** (2007).
95. Pelli, D.G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision* **10**, 437-442 (1997).
96. Breitmeyer, B.G. & Ogmen, H. Recent models and findings in visual backward masking: a comparison, review, and update. *Perception and Psychophysics* **62**, 1572-1595 (2000).
97. Acerbi, L. & Ma, W.J. Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Proceedings of Advances in Neural Information Processing Systems 30* (2017).