



# A dynamic normalization model of temporal attention

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

**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 modelled 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.**

The visual system receives continuous, time-varying input, which must be dynamically prioritized according to behavioural 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 recently developed 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<sup>1,2</sup>. Normalization appears to be widespread in both basic sensory<sup>2–6</sup> and higher-order perceptual and cognitive processing<sup>7–12</sup>. For this reason, it has been described as a “canonical cortical computation”<sup>1</sup>.

Several models of attention combine sensory normalization with attentional modulation<sup>6,12–19</sup>. 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<sup>12</sup>, proposes that attention modulates neural activity before normalization. This formulation has reconciled ostensibly conflicting electrophysiological and psychophysical findings<sup>12</sup> and predicted new results that have been empirically confirmed<sup>20–22</sup>. However, this leading theory of spatial and feature-based attention is static, with no temporal attention component.

Dynamic normalization models have been developed to account for the time courses of neuronal responses<sup>5,10,23–27</sup> and dynamic sensory processes such as adaptation<sup>2,28–32</sup>, but these models have not incorporated attention. It has been noted<sup>17</sup> that differential shunting equations can be used to implement normalization, as in shunting equation models of spatial attention<sup>33,34</sup>.

A major challenge in developing a dynamic normalization model of attention is establishing what the attentional gain dynamics actually are. The behavioural time courses of spatial attention have been characterized, revealing that voluntary spatial attention takes 300 ms to be allocated while involuntary spatial attention peaks at 90–120 ms (refs. <sup>35–38</sup>). However, visual attention is not only directed to locations in space; it is also directed to points in time.

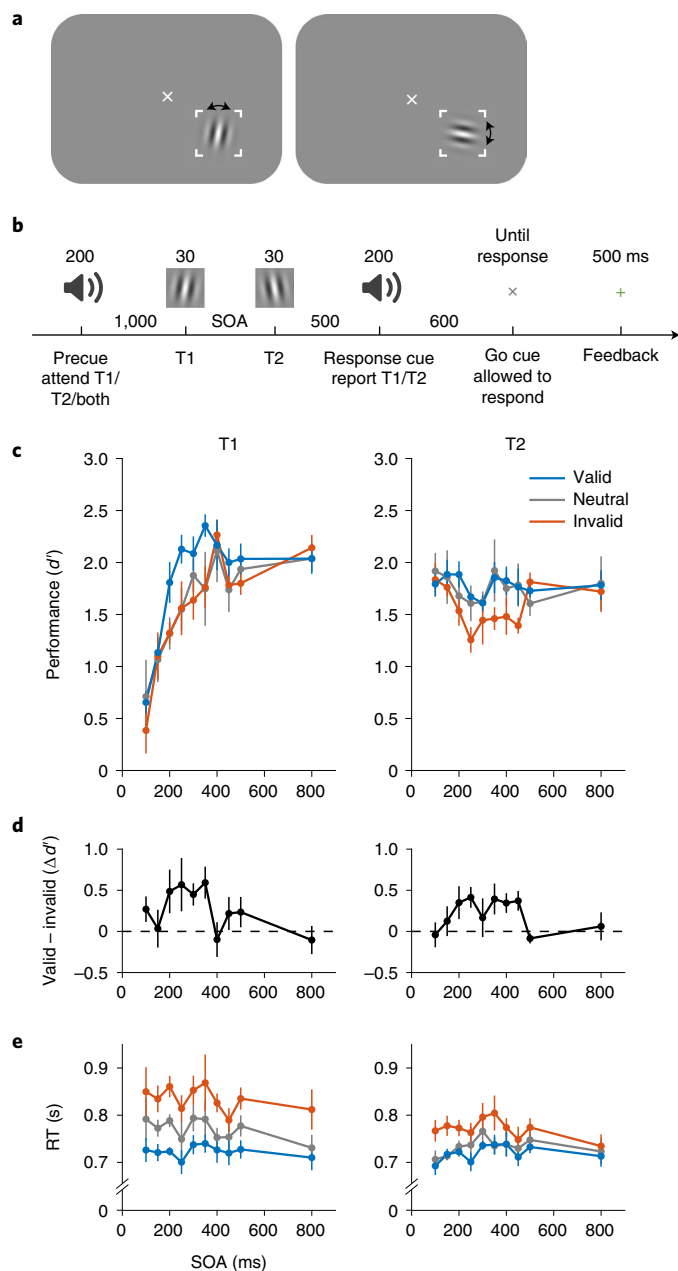
Temporal attention is the prioritization of visual information at specific points in time, for example, the moment a behaviourally relevant stimulus will appear<sup>39</sup>. Even with spatial attention fixed at one location, visual temporal attention can be manipulated using temporal precues to specific time points. Such voluntary, or goal-directed, temporal attention affects perception<sup>40–44</sup>, neural responses<sup>44–48</sup> and microsaccades<sup>49</sup>. 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<sup>41</sup>, 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 prioritizes specific points in time in an automatic 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<sup>35,38</sup>. However, the dynamics of involuntary temporal attention (even when spatial attention is fixed) are unknown, and there are no general-purpose models of involuntary temporal attentional dynamics.

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. We then used the model, with the same neuronal and attentional parameters, to fit two previous datasets<sup>41,42</sup>, thereby providing empirical evidence for the generalizability of our new model.

## Results

**Behaviour.** To determine the dynamics of voluntary and involuntary temporal attention, we performed a behavioural experiment (Fig. 1a,b). Observers judged the orientation of gratings while voluntary temporal attention was directed to different points in time.



**Fig. 1 | Behavioural protocol and data.** **a**, Stimulus display and task schematic. Observers discriminated CW versus 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 to 800 ms, but was fixed within each session. **c**, Perceptual sensitivity ( $d'$ ) for each SOA, precue (valid, neutral or invalid) and target interval (T1 or T2). Main effect of precue validity ( $F(2, 8) = 8.85, P = 0.0094, \eta_G^2 = 0.048$ ). **d**, Effect of precueing temporal attention on perceptual sensitivity (that is, difference in performance between valid and invalid precues). Size of precueing effect depended on SOA ( $F(9, 36) = 3.13, P = 0.0069, \eta_G^2 = 0.20$ ). **e**, RT. Error bars in **c** and **e** are within-observers s.e.m. computed separately for each target to reflect differences among conditions for each target. Error bars in **d** are s.e.m.,  $n = 5$ .

On each trial, two gratings appeared in sequence at the same location, separated by a stimulus onset asynchrony (SOA) that ranged from 100 to 800 ms across testing sessions but was fixed within a

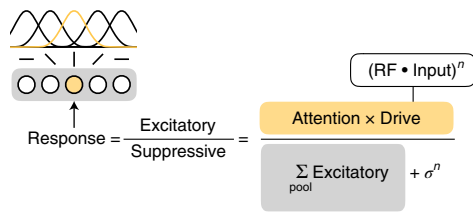
session to ensure predictable stimulus timing. Voluntary temporal attention was manipulated by an auditory precue to attend to the first target (T1), the second target (T2) or both targets (neutral precue). When a single target was precued (80% of all trials), precue validity was 75%. On valid attention trials (60% of all trials), 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. On neutral trials (20%), observers 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. <sup>41</sup>, 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 behaviour, as a function of SOA. Behavioural performance depended on the temporal attentional precue, the SOA and the reported target. We identified four main features of the behavioural data.

First, voluntary temporal attention affected behaviour, 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 (Fig. 1c,d). In a repeated-measures analysis of variance (ANOVA) with precue validity, SOA and target as factors, there was a main effect of validity ( $F(2, 8) = 8.85, P = 0.0094, \eta_G^2 = 0.048$ ). 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 (valid versus neutral) and costs (invalid versus neutral) separately for T1 and T2 yielded a marginally significant benefit for T1 ( $F(1, 4) = 5.00, P = 0.089, \eta_G^2 = 0.079$ ) but no evidence for a significant cost for T1 ( $F(1, 4) = 0.43, P = 0.55, \eta_G^2 = 0.0017$ ). Conversely, there was a significant cost for T2 ( $F(1, 4) = 15.10, P = 0.018, \eta_G^2 = 0.065$ ) but no evidence for a significant benefit for T2 ( $F(1, 4) = 0.089, P = 0.78, \eta_G^2 = 0.0015$ ). Reaction time (RT) showed a similar dependence on the attentional precue, with fastest responses for valid trials, slowest for invalid trials and intermediate responses for neutral trials (Fig. 1e; main effect of validity,  $F(2, 8) = 21.92, P < 0.001, \eta_G^2 = 0.27$ ), confirming that speed-accuracy tradeoffs did not drive the differences in  $d'$ . The presence and pattern of precueing effects indicates temporal attentional tradeoffs across time, consistent with our previous findings with a 250 ms SOA (ref. <sup>41</sup>).

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) but little or no effect at the shortest and longest SOAs. This SOA dependence can be seen in Fig. 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 ( $F(9, 36) = 3.13, P = 0.0069, \eta_G^2 = 0.20$ ). There was neither a main effect of target ( $F(1, 4) = 0.11, P = 0.76, \eta_G^2 = 0.0059$ ) nor an interaction between SOA and target ( $F(9, 36) = 1.00, P = 0.45, \eta_G^2 = 0.10$ ). The pattern of precueing effects was consistent across observers (Extended Data Fig. 1).

Third, the overall performance of T1 increased substantially with SOA, from  $d'$  of  $\sim 0.6$  at the 100 ms SOA to  $\sim 2.1$  at the 800 ms SOA (Fig. 1c), on average across precueing conditions (two-tailed paired  $t$  test,  $t(4) = 5.72, P = 0.0046$ , Cohen's  $d = 2.56$ , mean difference 1.49



**Fig. 2 | Static R&H normalization model of attention equation.** We consider a population of orientation-tuned neurons (left, grey-shaded background). The computation of the response of one of the neurons (shaded yellow, with corresponding yellow 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 grey background) plus a constant.

with 95% CI 0.77–2.21). We call this rising function of SOA for T1 “masking-like behaviour”<sup>50,51</sup>. The high T1 performance levels for the longest SOA (800 ms) 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 average point at 250 ms (Fig. 1c). The dip was larger for invalid trials (reaching  $d' = 1.3$  versus a maximum of 1.8) but was also present for valid and neutral trials ( $d' = 1.6$  versus a maximum of 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<sup>52,53</sup>, which has been much investigated both experimentally and through modelling<sup>52,54,55</sup>. 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 (ref. 56).

Statistically, the variation of  $d'$  across SOAs and targets was demonstrated by an effect of SOA on  $d'$  ( $F(9, 36) = 3.60, P = 0.0028, \eta_G^2 = 0.19$ ), which differed for the two targets (SOA  $\times$  target interaction,  $F(9, 36) = 15.38, P < 0.001, \eta_G^2 = 0.27$ ). For T2 specifically, the  $d'$  difference between 100 and 250 ms, on average across precueing conditions, was also significant (two-tailed paired  $t$  test,  $t(4) = 2.95, P = 0.042$ , Cohen's  $d = 1.32$ , mean difference 0.34 with 95% CI 0.020–0.65).

For RT (Fig. 1e), in addition to the main effect of validity, there was a trend toward faster T2 responses than T1 responses ( $F(1, 4) = 5.52, P = 0.078, \eta_G^2 = 0.091$ ), and the precue influenced RT less for T2 than for T1 (validity  $\times$  target interaction,  $F(2, 8) = 6.83, P = 0.019, \eta_G^2 = 0.048$ ). No other main effects or interactions were significant for  $d'$  or RT ( $F < 1.3$ ) (Extended Data Fig. 2).

In summary, the psychometric time courses for the two-target temporal precueing task were quite rich, with masking-like behaviour for T1, AB-like behaviour 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<sup>12</sup>

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<sup>12,15</sup> and has also been implicated in the effects of rhythmic expectation on perceptual sensitivity<sup>57–59</sup>.

Each layer of the model consists of a population of neurons whose responses follow the R&H equation (Fig. 2). Each neuron's response is determined by the same basic operations: bottom-up input to a neuron in a given layer is filtered through that neuron's receptive field (RF), multiplied by top-down attentional modulation, which we term ‘attentional gain’, and then divisively normalized by the activity of its neighbours.

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:

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

Here,  $r_i$  is the response of neuron  $i$  (where  $i \in \{1, \dots, N\}$  for a population of  $N$  neurons),  $e_i$  is the excitatory drive to the neuron,  $s_i$  is the suppressive drive,  $\sigma$  is a semi-saturation constant that prevents 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  $\tau$  is a time constant that determines how long the response takes to rise to steady state when the input turns on and to return to zero when it turns off.

The excitatory drive  $e_i$  was determined by the equation

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

where  $\mathbf{x}$  is the bottom-up input to the layer,  $\mathbf{w}_i$  is the RF of the neuron and  $a_i$  is the 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 the Methods section.

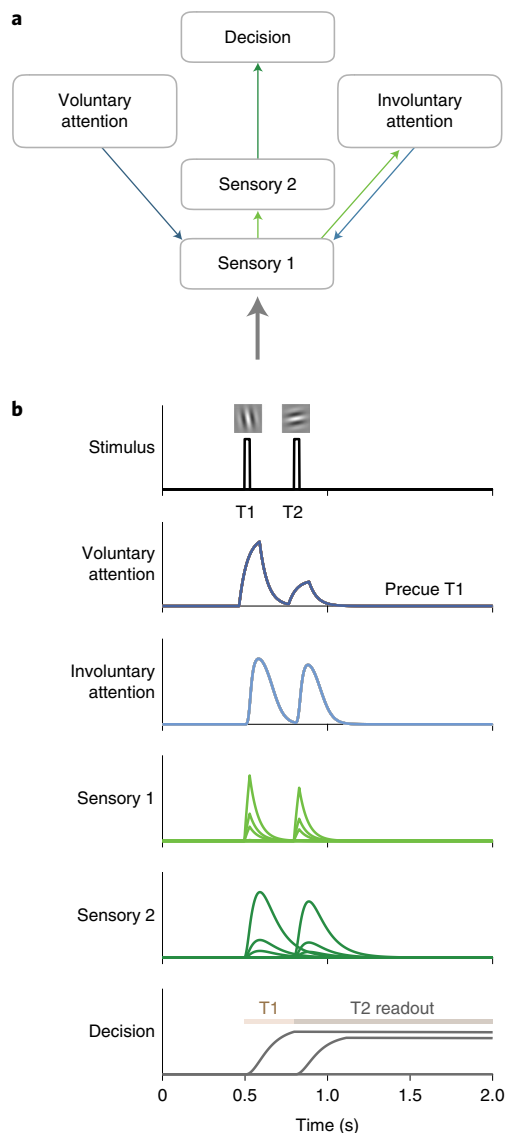
The suppressive drive  $s_i$  was determined by the equation

$$s_i = \sum_{j=1}^N e_j, \quad (3)$$

a summation of the excitatory drives of a pool of neurons. Here, the pool was simply all the neurons in the layer (for example, 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 the behavioural and neurophysiological effects of spatial and feature-based attention<sup>20,21</sup>, which have been successfully described and predicted by the R&H model<sup>12</sup>. Like the R&H model, this model is intended to be computationally clear but not biophysically precise; as discussed previously<sup>12</sup>, there are many biophysical mechanisms that could implement normalization. For example, a recently developed circuit model uses recurrence to implement normalization, with steady-state behaviour equivalent to the R&H equation<sup>27</sup>. In the current model, the ‘neurons’ should be thought of as mapping to computational units at the neural population level.

**Model specification.** The model architecture was a hierarchical, recurrent neural network, with sensory, attention and decision layers (Fig. 3a and Supplementary Table 1). The layers generated continuous neural response (firing rate) time series given attentional precues and continuous stimulus input (Fig. 3b). Each layer



**Fig. 3 | Model architecture and example simulated time series. a,**

Architecture, showing sensory input as a thick grey 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, with 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.

performed the same computation (equations 1–3); only the inputs and outputs were layer specific. Full details of the model and simulations can be found in the Methods section.

**Sensory layers.** The sensory layers represent visual cortical areas. Sensory layer 1 (S1) neurons are orientation selective and receive stimulus input: the stimulus orientation (or zero for no stimulus) at every point in time. It also receives top-down attentional modulation from both the voluntary and involuntary attention layers. Voluntary and involuntary attention combine multiplicatively to determine the attentional gain  $a$  for S1. Sensory layer 2 (S2) receives

input from S1, inheriting its orientation tuning. S2 has a slower rise and more sustained responses than S1 (because the input to S2 is the output from S1), which helps capture T1 behavioural performance as a function of SOA.

**Voluntary attention layer.** The voluntary attention layer (VA) increases attentional gain at task-relevant times. Responses in VA depend on the precue (T1, T2 or neutral) and the trial timing. The input to VA is a time-varying control signal (Fig. 4) that reflects the observer’s knowledge of the precue and SOA. The control signal consists of square wave pulses around the times of each target. Pulse latency and duration are free parameters. The amplitude of each pulse is determined by the allocation of voluntary attention to each target (that is, more voluntary attention at a certain time point generates a larger pulse at that time). These control pulses in turn determine the VA response and corresponding attentional gain modulation of S1.

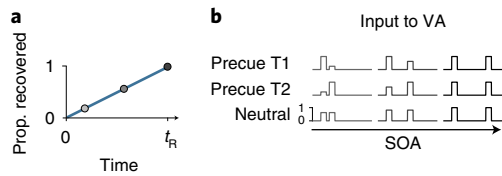
Voluntary attention was a limited resource across time, generalizing the idea of limited spatial attention resources (see, for example, refs. 21,35,60–63) to the time domain (see Modelling the data section). Immediately after a maximum ( $=1$ ) allocation of attention, none was available, but attention recovered over time (Fig. 4a). We modelled 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 (Fig. 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 governed the attentional allocation.

**Involuntary attention layer.** The involuntary attention layer (IA) is stimulus driven, receiving input from S1. It also feeds back to S1, providing a second source of attentional modulation. Because IA responses are driven by S1, they started slightly later than S1 responses (Fig. 3b). Further, their magnitude depends on the voluntary attentional modulation of S1, because larger S1 responses drive larger IA responses (Fig. 3b).

**Decision layer.** The decision layer (D) represents a decision area (for example, in parietal cortex<sup>64</sup>) and receives input from S2. An optimal linear classifier is used to decode clockwise (CW) versus counterclockwise (CCW) evidence at each time step from the S2 population. This decoded sensory evidence is the input drive to D. The time constant for D is fixed to be long, which allows D to accumulate sensory evidence over time, without leakage, similar to drift–diffusion models<sup>65</sup>. Decision neurons were target specific, accumulating evidence only during a corresponding target readout window (Fig. 3b). The model’s task performance was determined by the response of the decision neuron representing the target that was cued (by the response cue) at the end of the trial.

**Modelling the data. Main model.** The normalization model of dynamic attention fitted 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 behaviour for T1 and (4) AB-like behaviour for T2 (Fig. 5a). Fitted parameter values are listed in Table 1.

To capture the two behavioural features related to voluntary temporal attention—that is, 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 (Fig. 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



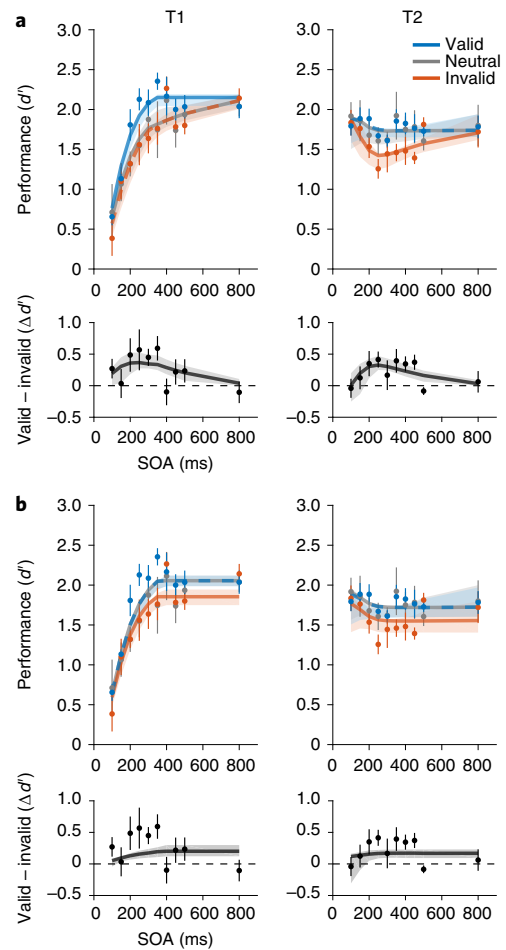
**Fig. 4 | Voluntary attention as a limited but recoverable resource. a,**

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 types 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, equal attention was allocated to T1 and T2 in the neutral condition.

neutral condition<sup>60,61,63</sup>. 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. However, 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 Supplementary Tables 2 and 3.

The overall shapes of the performance functions for T1 and T2 are produced by additional model components. The masking-like behaviour for T1 is produced by stopping the decision readout for T1 when T2 appears. The AB-like behaviour for T2 is produced by a combination of three factors: (1) Limited voluntary attention results in lower performance at shorter SOAs, especially for invalid trials. (2) At the shortest SOAs (~100 ms), voluntary attention to T1 is sustained long enough to enhance both T1 and T2 sensory responses, boosting T2 performance. (3) Involuntary attentional excitation combines 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 fitted the data almost as well ( $R^2 = 0.89$ ) and was better in model comparison due to having fewer parameters (difference in Akaike information criterion ( $\Delta\text{AIC}$ ) = -5.8), although it could not produce AB-like behaviour for T2 valid trials (Supplementary Results).

**No limit variant.** A model without limited voluntary attention (Fig. 5b) produced a poorer fit ( $R^2 = 0.83$ ,  $\Delta\text{AIC} = 26$  with respect to the main model). It also failed to capture the data qualitatively (Fig. 5b) in two ways: (1) The no limit variant did not produce tradeoffs in temporal precueing effects. It predicted that neutral performance was equal to valid performance for both T1 and T2, unlike in the data, where T1 neutral performance was similar to invalid performance. (2) The no limit variant did not produce peak precueing effects at intermediate SOAs. Rather, the longest SOAs had maximal precueing effects. These failures of the model are due to its structure and could not be altered by a different choice of parameters. A model recovery analysis confirmed the distinguishability of the no limit variant from the main model variant (Supplementary Fig. 1).



**Fig. 5 | Model fits to perceptual sensitivity data. a,** Main model with a voluntary attention limit ( $R^2 = 0.90$ ): performance ( $d'$ , top row) and precueing effect (bottom row). **b,** No limit model variant ( $R^2 = 0.83$ ). The main model fitted better than the no limit model variant ( $\Delta\text{AIC} = -26$ ). Behavioural performance (data points and error bars, copied from Fig. 1,  $n = 5$ ) and model fits (curved lines). Some curves are dashed to reveal overlapping model predictions. Shaded regions show bootstrapped 95% CIs for model fits.

The performance of the no limit model variant reveals why a limit on voluntary attention is necessary. The fact that neutral performance is very similar to valid performance for both targets shows that, without the limit, the model has 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 a difference between valid and invalid performance into this model variant 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 trade-off incentives without a limit on voluntary attention (Supplementary Fig. 2 and Supplementary Tables 1 and 2).

**Table 1 | Model parameters**

Parameter	Description	Value	95% CI
<b>All layers</b>			
$n$	Exponent	1.5	—
<b>Sensory layer 1</b>			
$\tau_{S1}$	Time constant	52	49–116
$\sigma_{S1}$	Semi-saturation constant	1.4	1.2–2.0
<b>Sensory layer 2</b>			
$\tau_{S2}$	Time constant	100	69–120
$\sigma_{S2}$	Semi-saturation constant	0.1	—
<b>Decision layer</b>			
$\tau_D$	Time constant	1e5	—
$\sigma_D$	Semi-saturation constant	0.7	—
<b>Voluntary attention layer</b>			
$\tau_{VA}$	Time constant	50	—
$\sigma_A$ (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 to –6
$t_{VADur}$	Duration of voluntary control signal	124	99–374
$t_R$	Recovery time of voluntary gain	918	600–1,091
$w_N$	Weight to treat neutral precue more like precue T1 (1) or precue T2 (0)	0.28	0.01–0.53
<b>Involuntary attention layer</b>			
$\tau_{IA}$	Time constant	2	—
$\sigma_A$ (shared with VA)	Semi-saturation constant	20	—
$b_{IA}$	Amplitude of involuntary gain modulation	8.5	0.8–27.9
$h_{IA;p}$	Shape parameter for involuntary temporal prefilter	2.2	0.04–49.9
$h_{IA;q}$	Scaling parameter for involuntary temporal prefilter	0.023	0.01–0.09
<b>Fitting</b>			
$s_{T1}$	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
<b>Number of parameters</b>			
Total		20	
Fitted		12	

Parameters listed without CIs were fixed to set values and not optimized during fitting. All times (that is, time constants and latencies) are given in milliseconds. Negative latencies for  $t_{VAOn}$  mean that the voluntary control signal started before the target.

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 fitted to the data, neither of these implementations produced trade-off profiles. We found that limited voluntary attention was still required to let each of these model

variants fit the data (Supplementary Fig. 3 and Supplementary Table 4), with  $t_R$  estimates of 809 and 924 ms, respectively.

**Generalization to independent datasets and other tasks.** To test the ability of the main model to generalize to independent datasets, we fitted the model to data from two previous experiments that used the same voluntary temporal attention task<sup>41,42</sup>. To do so, we fixed all parameters to the best-fit values from the current experiment and fit only two free parameters to each dataset to scale the overall performance of T1 and T2. Thus, the relative magnitudes of the T1 and T2 attention effects and the tradeoffs between benefits to one target (valid versus neutral) and costs to the other (invalid versus neutral) were fixed to the values in Table 1. The model fitted the new data reasonably well, with  $R^2 = 0.83$  for the full datasets from each experiment (Fig. 6a,b). Fits to separate conditions in ref. 42, in which the stimulus was placed at different visual field locations, yielded  $R^2$  values of 0.73 to 0.95 (Fig. 6c). The model slightly underestimated the precueing effect size for T1 in ref. 41. However, it correctly predicted the smaller precueing effect for T2 compared with T1 in ref. 42, due to the biased attentional trade-off between targets on neutral precue trials, controlled by  $w_N$ . Thus, with parameters for all neuronal and attentional dynamics fixed, the current model could capture independent datasets.

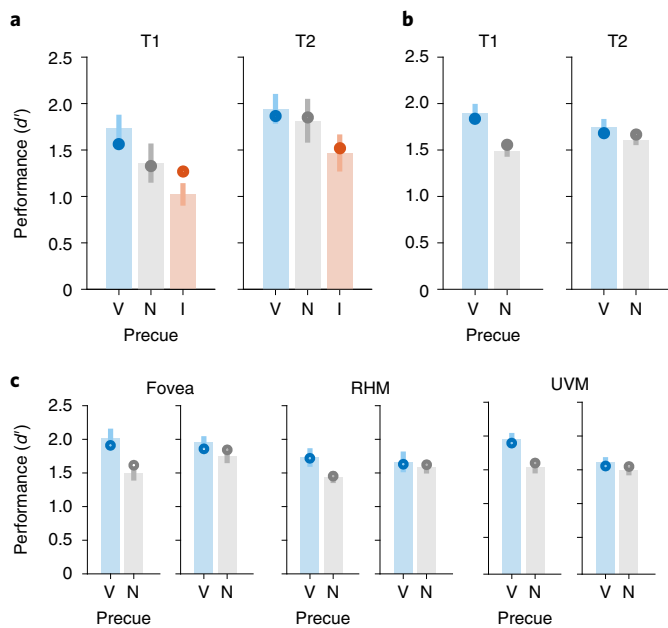
Finally, we asked whether our modelling framework could capture performance in an AB task. The AB task differs from the two-target temporal cueing task in multiple ways, including (1) targets are embedded in a rapid sequence of non-target stimuli and (2) target timing is unpredictable, so voluntary temporal attention cannot be allocated in advance of the targets. A preliminary simulation showed that the current model can capture major features of the AB (Supplementary Fig. 4) and invites further work to test the normalization model of dynamic attention on the AB and other tasks.

## Discussion

**A normalization model of dynamic attention.** We developed a normalization model of dynamic attention, generalizing the R&H normalization model of attention<sup>12</sup> 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, that is, 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 behaviour for T1 and AB-like behaviour for T2.

The model could reproduce the behavioural 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<sup>38,63,66–70</sup>. Although incorporating involuntary attentional gain dynamics into our modelling framework was theoretically motivated, model comparison showed that it was not required to fit the current psychophysical data quantitatively. Future research will be needed to further examine whether and how involuntary attention interacts with voluntary temporal attention.



**Fig. 6 | Generalization to independent datasets.** **a–c**, The main model was fitted (two free parameters, all others fixed) to previously published datasets: **(a)** experiment 1 of ref. <sup>41</sup> ( $n = 10$ ,  $R^2 = 0.83$ ); and ref. <sup>42</sup> **(b)** averaged across all conditions ( $n = 10$ ,  $R^2 = 0.83$ ) and **(c)** separated by visual field location condition ( $R^2 = 0.73$ – $0.95$ ) for fovea, right horizontal meridian (RHM) and upper vertical meridian (UVM). In both experiments, the two targets were separated by a 250 ms SOA. Bars and error bars show data with s.e.m. Points show model fits. V, valid; N, neutral; I, invalid.

Voluntary attention took the form of a limited resource that recovered over time, with full recovery estimated to take  $\sim 1$  s, and consistent with a range of 0.6–1.1 s (95% CI). The attentional limitation in our model could be, for example, either a limitation on the available voluntary attentional gain ( $r^{VA}$ ) 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 behaviour. It therefore makes specific predictions about different types of attentional gain dynamics (voluntary and involuntary attention time courses). Alternative model variants also required a limitation on voluntary attention across time, but they predicted different gain dynamics. 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<sup>35,60,61</sup> and working memory<sup>71</sup>. Here, we propose a limited resource across time that underlies the selectivity of voluntary attention to points in time.

**Relation to other attention models concerned with dynamics.** Previous modelling frameworks that incorporate both attention and some dynamic element include: the attention gating model<sup>72–74</sup>, the theory of visual attention (TVA)<sup>75–77</sup> and the competitive interaction theory<sup>17,33,34</sup>. Each framework includes different model variants, some of which incorporate normalization<sup>17,75</sup>. These models have had success in accounting for behavioural data from various perceptual tasks. Other frameworks focus on rhythmic attention<sup>78,79</sup>, which we do not consider here.

There are several important differences between these models and our dynamic attention model. First, we model voluntary temporal attention. TVA has been adapted to model a constant level of

expectation across time<sup>80,81</sup>, 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<sup>35,82,83</sup> and has been reported for temporal attention<sup>78,84–86</sup>. Third, in previous models<sup>17,77</sup>, 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 working-memory encoding. Fourth, our model is built to handle time-varying stimuli and time-varying attention, rather than single, brief displays<sup>33</sup>, and without being constrained by attentional episodes<sup>73</sup>. 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 AB?** T2 performance in our two-target temporal precue task resembled T2 performance in AB tasks<sup>52</sup>. The fact that we observed AB-like behaviour 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 behaviour<sup>87,88</sup>. The few AB studies in which voluntary temporal attention has been manipulated have reported inconsistent findings<sup>89–91</sup>. Here, we manipulated voluntary temporal attention and tested different model variants in which voluntary attentional dynamics either contributed to or were independent from AB-like behaviour. In our main model, the AB arises from limited voluntary attention. We found no need to invoke other processes previously proposed to contribute to the AB (for example, working memory limitations, loss of top-down control, and alpha oscillations)<sup>54,92,93</sup> to explain the AB-like behaviour in our task. However, the contribution of such processes is not excluded by our model, and as yet we have no evidence that our model should be preferred over others to explain the AB per se.

In an influential AB model<sup>55</sup>, attention is enhanced by the appearance of a target and suppressed during working memory encoding, which leads to the AB. 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<sup>94</sup>. 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<sup>95</sup>. As the goal of the current study is to investigate voluntary temporal attention and not the AB, future work will be required to compare alternative models on a variety of tasks in which dynamic attention has been implicated, including the AB task. As a first step, we simulated an AB task and found that our model captures the major features of the AB.

**Future extensions of the model.** The current model is a general description of the dynamic interactions between attention and sensory responses. We focus on how attention affects sensory processing of oriented gratings, a strategy that has proven productive in studies of spatial attention<sup>12,15,21,35,61,63,68,82,83,96</sup>, facilitated by our knowledge of how orientation is represented in the visual system<sup>97</sup>. Future work could extend the model to include working memory layers, as sequential processing limitations may also arise in working memory<sup>54</sup>, as well as more complex feature representations (using additional sensory layers and different RFs) to handle more complex stimuli. It could also investigate how different types of noise at different stages of the model impact model behaviour. Here the limited resource of voluntary attention was implemented via constraints on the amplitude of the attention control signal over time, resulting in limits on attentional gain. Future work should

explore other implementations of a limited resource on voluntary attention over time.

## Conclusion

We developed a model of voluntary and involuntary visual temporal attention, which can serve as a general-purpose computational framework for modelling 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, behaviourally relevant timescales. The model predicts that voluntary temporal 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

**Behaviour.** *Observers.* Five human observers (20–30 years old, three female and two male) participated in the experiment. All observers provided informed consent, and the University Committee on Activities Involving Human Subjects at New York University (NYU) 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. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous psychophysics publications that took a similar approach of collecting a large amount of data per observer (for example, ref. 7).

*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<sup>98–100</sup> 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 four cycles per degree of visual angle sinusoidal gratings with a two-dimensional Gaussian spatial envelope (standard deviation 0.7°), presented in the lower-right quadrant of the display centred at 5.7° eccentricity (Fig. 1a). (The stimulus was placed in this quadrant in anticipation of future neuroimaging studies. We have previously shown the effects of voluntary temporal attention on orientation discrimination to be indistinguishable at different iso-eccentric peripheral locations and the fovea, so we expect stimulus location should not impact the results<sup>42</sup>.) Stimulus contrast was 64%. Placeholders, corners of a 4.25° x 4.25° white square outline (line width 0.08°) centred on the target location, were present throughout the display to minimize spatial uncertainty. The stimuli were presented on a medium-grey background (57 cd m<sup>-2</sup>). Auditory precues were high-frequency (784 Hz; G5) or low-frequency (523 Hz; C5) 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 interact dynamically to affect perception. We used the two-target temporal precueing task developed and previously described by Denison et al.<sup>41,49</sup>

*Task.* Observers discriminated the orientation of grating patches (Fig. 1). On each trial, two targets (T1 and T2) were presented at the same spatial location, separated by a fixed 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 CW or 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, such as 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 the 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 1,000 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 20 testing sessions (ten SOAs × two sessions per SOA, 9,600 trials in total), and two observers completed ten sessions (ten SOAs × one session per SOA, 4,800 trials in total) on separate days. The SOA order was randomly determined for each observer. Observers who completed two sessions per SOA did two sets of ten, with a separate random shuffling for each set. Each session consisted of all combinations of precue type (60% valid, 20% invalid and 20% neutral), probed target (T1 and T2), target tilt (CW and CCW, independent for T1 and T2) and target axis (horizontal and 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 three-down one-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 ten sessions, the tilt was adjusted before the second set of ten sessions.)

*Eye tracking.* Eye position was recorded using an EyeLink 1000 eye tracker (SR Research) at a sampling rate of 1,000 Hz. Raw gaze positions were converted into degrees of visual angle using the five-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 behaviour, we conducted repeated-measures ANOVAs and calculated  $\eta_p^2$  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, but note that reaction time here was a secondary measure that was analysed 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. Cohen's  $d$  for paired  $t$  tests was calculated as the mean of the paired differences divided by the standard deviation of the paired differences.

*Model.* The variables  $e$ ,  $s$ ,  $r$ ,  $a$ ,  $g$ ,  $x$ ,  $y$  and  $z$  are all time varying, for example,  $e(t)$ , but for simplicity of notation, we omit the time  $t$ . 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, for example,  $e_i^{s1}$  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 consists of a hierarchical, recurrent neural network, with different layers (Fig. 3a): sensory layers, analogous to visual cortical areas; attention layers, which modulate the sensory responses; and a decision layer, which reads out the sensory responses and reports a task decision (here, CW versus CCW grating tilt). All model parameters are listed in Table 1.

**Sensory layer 1.** Layer S1 (Fig. 3b) represents an early-stage visual area and receives stimulus input at every time step. The excitatory drive for each neuron  $i$  is

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

The stimulus  $\mathbf{x}$  is represented in terms of its orientation: a vector of length  $M$ , with each element corresponding to a different orientation  $\theta$ . When the target grating is on, the element corresponding to the stimulus orientation has a value  $c$ , the stimulus contrast:  $\mathbf{x} = (0, 0, \dots, c, \dots, 0)$ . When the grating is off, all elements are zero-valued. We showed the model the same two-target trial sequences (Fig. 3b, Stimulus) as we showed observers (Fig. 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  $\mathbf{w}_i^{S1} = (w_{i,1}, w_{i,2}, \dots, w_{i,M})$  was designed to be an orientation tuning curve described by one cycle of a raised cosine:

$$w_{ij} = |\cos(\theta_j - \varphi_i)^m|, \quad (5)$$

where  $\theta$  is orientation, with sampled orientations indexed by  $j$ , and  $\varphi_i$  is the preferred orientation of the  $i$ th neuron. Preferred orientations were evenly spaced,  $\varphi = (0, \pi/N, \dots, \pi - \pi/N)$ . 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 VA and IA, whose responses independently and multiplicatively modulated the sensory drive. The attentional gain  $a_i^{S1}$  is

$$a_i^{S1} = \left[ 1 + b_{VA} r_i^{VA} \right] \left[ 1 + b_{IA} r_i^{IA} \right], \quad (6)$$

where  $r^{VA}$  and  $r^{IA}$  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 half-wave 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 1, sensory responses increased above baseline; we call this ‘excitatory’ attentional modulation. When  $a$  was less than 1, sensory responses decreased below baseline; we call this ‘inhibitory’ attentional modulation.

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

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

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  $\tau_{S2}$ . Therefore, S2 had a slower rise and more sustained responses than S1, which helped capture T1 behavioural performance as a function of SOA. However, temporal RFs with more complex dynamics within a layer (for example, a cascade of exponentials) could have achieved the same result using only one sensory layer.

**Decision layer.** The decision layer (Fig. 3b) represents a decision area (for example, in parietal cortex<sup>43</sup>) and receives input from S2. To encode information about temporal order, there are two neurons in the decision layer, one for T1 and one for T2. A binary (0 or 1) decision window  $g$  gates the input to each neuron: a neuron receives input drive only when its decision window is on. The T1 neuron’s decision window starts at the onset of T1 and stops at the onset of T2. The T2 neuron’s decision window starts at the onset of T2 and stops at the end of the trial. Thus, the decision layer reads out from successive decision windows for the two targets (Fig. 3b, Decision, shaded regions). The decision window cutoff for T1 implements a simplified version of masking, standing in for a mechanism that curtails T1-related signals when T2 appears<sup>50,101</sup>.

The input drive to the decision layer is the evidence for CW or CCW tilts, based on the responses of S2. The excitatory drive for each neuron  $i$  is

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

The inner product in this equation represents an optimal linear classifier that decodes CW versus CCW evidence at each time step from the S2 population response. The classifier projects  $\mathbf{r}^{S2}$  onto the difference  $\mathbf{w}^D$  between two templates corresponding to ideal responses to the two possible stimuli (CW and CCW). These templates are ‘ideal’ in the sense that they are the population responses for CW or CCW stimuli at full contrast and with no noise. (We assume that the orientation axis—vertical or horizontal—is known to the observer, so the comparison is only between the CW and CCW templates on the relevant orientation axis.) The classifier projection gives a continuous value that indicates the similarity of the population response on that trial to the CW versus CCW templates. We arbitrarily assign 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 indicates a CW versus CCW choice, and the magnitude indicates the strength of evidence for the choice. Because sensory evidence at each time step creates a new input drive to the decision layer, the decision layer accumulates evidence across time, similar to drift-diffusion models<sup>65</sup>. We fix 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 is used to determine the model’s performance. Specifically, the model’s CW versus CCW choice on each trial is determined by the sign of the response, and the magnitude of the decision neuron’s response is 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 fitted 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<sup>51,42</sup>, but we let these differences be captured by the T2 scaling parameter.

**Voluntary attention layer.** Responses in VA (Fig. 3b) depend on the precue for a given trial (for example, precue T1) and knowledge of task timing, so that VA responses increase at the time of the relevant sensory responses. The input  $y$  to the layer is a time-varying control signal that reflects the observer’s knowledge of the SOA and the precue type (T1, T2 or neutral). The excitatory drive is determined by the input as

$$e_i^{VA} = y^n. \quad (9)$$

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

In the main model variant, voluntary attentional gain was a limited resource across time. The amplitudes of the square waves then depend 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 the available gain would recover to the maximum level over time (Fig. 4). We modelled 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

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

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 1,000 ms and the SOA was 400 ms, the total attention available across both targets was 1 (to the precued target) + 400/1,000 (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. Figure 4 shows examples of the attention control input to VA. We chose the linear recovery function for simplicity and interpretability, as it gives rise to attentional tradeoffs between targets that depend only on  $t_R$  and  $w_N$ . Other types of recovery functions, such as exponential recovery, could be explored in future work.

**Involuntary attention layer.** Responses in IA are driven by input from S1 (that is, 'stimulus driven'), as shown in Fig. 3. The excitatory drive is

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

where  $\mathbf{z}$  is a temporally filtered version of  $\mathbf{r}^{S1}$ . The temporal filter, which we refer to as a 'prefilter', is 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  $\mathbf{w}^{IA}$  is the same for all the neurons in the layer and weights all S1 responses equally, that is, was not feature selective. Because IA responses are driven by S1, they start slightly later than S1 responses (Fig. 3b). Further, their magnitude depends on the voluntary attentional modulation of S1, with stronger involuntary responses to sensory responses that are voluntarily attended (Fig. 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 behavioural 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:  $t_{p1}$  and  $w_{S1}$ .

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, to T2 or to both targets (neutral precue). So, for precue T1 trials, the model allocates 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 allocates 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 behavioural data without a temporal limit on voluntary attention. These model variants are described in the [Supplementary Results](#) section.

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

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

To generate the model performance for one condition (defined by a combination of precue and SOA), a single trial was simulated recursively for 1,050 time points  $t$ . 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  $\times$  10 SOAs  $\times$  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 behaviour 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 (the sum of the squared error between the model output and data) was evaluated at 2,000 parameter sets sampled from reasonable parameter ranges, which were the same for all models tested. To sample evenly across the full range, each range was divided into 400 bins, and five 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 (BADSD)<sup>102</sup>, 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 (that is, not fit) for theoretical reasons or to minimize redundancies among the fitted parameters. Fixed parameters are those listed without confidence intervals (CIs) in Table 1. The values of all fixed and best-fit parameters are given in Table 1.

**Resampling procedures.** We obtained CIs 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. CIs on parameter estimates (Table 1) were calculated from the bootstrapped estimates. CIs on model fits (Fig. 5) were calculated for each condition combination 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, the onset and duration of the voluntary attention control signal, the temporal prefilter for the involuntary attention layer and the amplitudes of the 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 with the other amplitudes from a given fit.

**Model comparison.** We compared models using the AIC, computed with the assumption of a normal error distribution<sup>103</sup>,

$$\text{AIC} = n \ln \left( \frac{\text{RSS}}{n} \right) + 2k + C, \quad (13)$$

where  $n$  is the number of observations, RSS is the residual sum of squares,  $k$  is the number of free parameters and  $C$  is a constant that is the same across models. To compare models, we computed  $\Delta\text{AIC}$ . Models with smaller AIC values fit the data better.

**Generalization to independent datasets.** We fitted the main model to data from two previous experiments<sup>41,42</sup>. In ref. 41, discrimination performance for experiment 1 was reported as percent correct, so we reanalysed the data to calculate  $d'$  from trials aggregated for each precue type. In ref. 42, the data from each visual field location were fitted separately, and because there were no significant differences across locations in that study, we also fitted the average data across locations. For each fit, two parameters were allowed to vary:  $s_{T1}$  and  $s_{T2}$ . These parameters are scaling constants that control the overall performance of each target. All other parameters were fixed to the best-fitting values from the main model fit to the current data (Table 1). The fitting procedures were otherwise the same as for the current data. Procedures for the AB simulation are described in the [Supplementary Results](#) section.

**Reporting summary.** Further information on research design is available in the [Nature Research Reporting Summary](#) linked to this article.

## Data availability

All behavioural data are publicly available on the Open Science Framework (OSF) (<https://osf.io/dkx7n>).

## Code availability

All custom code for the model is publicly available on OSF (<https://osf.io/dkx7n>). Code for the behavioural experiments is available on GitHub (<https://github.com/racheldenison/temporal-attention>).

Received: 15 January 2020; Accepted: 29 April 2021;  
Published online: 17 June 2021

## References

- Carandini, M. & Heeger, D. J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2012).
- Heeger, D. J. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992).
- Bonin, V., Mante, V. & Carandini, M. The suppressive field of neurons in lateral geniculate nucleus. *J. Neurosci.* **25**, 10844–10856 (2005).
- 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. *J. Neurophysiol.* **118**, 1903–1913 (2017).
7. Li, H.-H., Carrasco, M. & Heeger, D. J. Deconstructing interocular suppression: attention and divisive normalization. *PLoS Comput. Biol.* **11**, e1004510 (2015).
8. Li, H.-H., Rankin, J., Rinzel, J., Carrasco, M. & Heeger, D. J. Attention model of binocular rivalry. *Proc. Natl Acad. Sci. USA* **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. *J. Neurosci.* **34**, 16046–16057 (2014).
11. Ohshiro, T., Angelaki, D. E. & Deangelis, G. C. A normalization model of multisensory integration. *Nat. Neurosci.* **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. *Vis. Res.* **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. *Annu. Rev. Vis. Sci.* **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 Comput. Biol.* **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. *Vis. Res.* **116**, 219–240 (2015).
18. Ni, A. M. & Maunsell, J. H. R. Neuronal effects of spatial and feature attention differ due to normalization. *J. Neurosci.* **39**, 5493–5505 (2019).
19. Beuth, F. & Hamker, F. H. A mechanistic cortical microcircuit of attention for amplification, normalization and suppression. *Vis. Res.* **116**, 241–257 (2015).
20. Herrmann, K., Heeger, D. J. & Carrasco, M. Feature-based attention enhances performance by increasing response gain. *Vis. Res.* **74**, 10–20 (2012).
21. Herrmann, K., Montaser-Kouhsari, L., Carrasco, M. & Heeger, D. J. When size matters: attention affects performance by contrast or response gain. *Nat. Neurosci.* **13**, 1554–1559 (2010).
22. Zhang, X., Japee, S., Safiullah, Z., Mlynaryk, N. & Ungerleider, L. G. A normalization framework for emotional attention. *PLoS Biol.* **14**, e1002578 (2016).
23. Carandini, M., Heeger, D. J. & Movshon, J. A. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17**, 8621–8644 (1997).
24. Reynaud, A., Masson, G. S. & Chavane, F. Dynamics of local input normalization result from balanced short- and long-range intracortical interactions in area V1. *J. Neurosci.* **32**, 12558–12569 (2012).
25. Sit, Y. F., Chen, Y., Geisler, W. S., Miiikkulainen, R. & Seidemann, E. Complex dynamics of V1 population responses explained by a simple gain-control model. *Neuron* **64**, 943–956 (2009).
26. Zhou, J., Benson, N. C., Kay, K. N. & Winawer, J. Compressive temporal summation in human visual cortex. *J. Neurosci.* **38**, 691–709 (2018).
27. Heeger, D. J. & Zemlianova, K. O. A recurrent circuit implements normalization, simulating the dynamics of V1 activity. *Proc. Natl Acad. Sci. U. S. A.* **117**, 22494–22505 (2020).
28. Wainwright, M. J., Schwartz, O. & Simoncelli, E. P. in *Statistical Theories of the Brain* (eds Rao, R. P. et al.) 1–22 (MIT Press, 2002).
29. Westrick, Z. M., Heeger, D. J. & Landy, M. S. Pattern adaptation and normalization reweighting. *J. Neurosci.* **36**, 9805–9816 (2016).
30. Wilson, H. R. & Humanski, R. Spatial frequency adaptation and contrast gain control. *Vis. Res.* **33**, 1133–1149 (1993).
31. Wissig, S. C. & Kohn, A. The influence of surround suppression on adaptation effects in primary visual cortex. *J. Neurophysiol.* **107**, 3370–3384 (2012).
32. Kalikhovich, D. A. & Vogels, R. Divisive normalization predicts adaptation-induced response changes in macaque inferior temporal cortex. *J. Neurosci.* **36**, 6116–6128 (2016).
33. Smith, P. L. & Sewell, D. K. A competitive interaction theory of attentional selection and decision making in brief, multielement displays. *Psychol. Rev.* **120**, 589–627 (2013).
34. Smith, P. L. & Ratcliff, R. An integrated theory of attention and decision making in visual signal detection. *Psychol. Rev.* **116**, 283–317 (2009).
35. Carrasco, M. Visual attention: the past 25 years. *Vis. Res.* **51**, 1484–1525 (2011).
36. Carrasco, M., Ling, S. & Read, S. Attention alters appearance. *Nat. Neurosci.* **7**, 308–313 (2004).
37. Liu, T., Stevens, S. T. & Carrasco, M. Comparing the time course and efficacy of spatial and feature-based attention. *Vis. Res.* **47**, 108–113 (2007).
38. Müller, H. J. & Rabbitt, P. M. Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 315–330 (1989).
39. Nobre, A. C. & van Ede, F. Anticipated moments: temporal structure in attention. *Nat. Rev. Neurosci.* **19**, 34–48 (2018).
40. Correa, A., Lupiáñez, J. & Tudela, P. Attentional preparation based on temporal expectancy modulates processing at the perceptual level. *Psychon. Bull. Rev.* **12**, 328–334 (2005).
41. Denison, R. N., Heeger, D. J. & Carrasco, M. Attention flexibly trades off across points in time. *Psychon. Bull. Rev.* **24**, 1142–1151 (2017).
42. Fernández, A., Denison, R. N. & Carrasco, M. Temporal attention improves perception similarly at foveal and parafoveal locations. *J. Vis.* **19**, 12 (2019).
43. Rohenkohl, G., Gould, I. C., Pessoa, J. & Nobre, A. C. Combining spatial and temporal expectations to improve visual perception. *J. Vis.* **14**, 8 (2014).
44. 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. *Proc. Natl Acad. Sci. USA* **112**, 8439–8444 (2015).
45. Anderson, B. & Sheinberg, D. L. Effects of temporal context and temporal expectancy on neural activity in inferior temporal cortex. *Neuropsychologia* **46**, 947–957 (2008).
46. Correa, A., Lupiáñez, J., Madrid, E. & Tudela, P. Temporal attention enhances early visual processing: a review and new evidence from event-related potentials. *Brain Res.* **1076**, 116–128 (2006).
47. 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. *J. Neurosci.* **18**, 7426–7435 (1998).
48. Miniussi, C., Wilding, E. L., Coull, J. T. & Nobre, A. C. Orienting attention in time. Modulation of brain potentials. *Brain* **122**, 1507–1518 (1999).
49. Denison, R. N., Yuval-Greenberg, S. & Carrasco, M. Directing voluntary temporal attention increases fixational stability. *J. Neurosci.* **39**, 353–363 (2019).
50. Breitmeyer, B. & Ogmen, H. *Visual Masking* (Oxford Univ. Press, 2006).
51. Kahneman, D. Method, findings, and theory in studies of visual masking. *Psychol. Bull.* **70**, 404–425 (1968).
52. Dux, P. E. & Marois, R. The attentional blink: a review of data and theory. *Atten. Percept. Psychophys.* **71**, 1683–1700 (2009).
53. Raymond, J. E., Shapiro, K. L. & Arnell, K. M. Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 849–860 (1992).
54. Chun, M. M. & Potter, M. C. A two-stage model for multiple target detection in rapid serial visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* **21**, 109–127 (1995).
55. Wyble, B., Potter, M. C., Bowman, H. & Nieuwenstein, M. Attentional episodes in visual perception. *J. Exp. Psychol. Gen.* **140**, 488–505 (2011).
56. 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. *J. Exp. Psychol. Learn. Mem. Cogn.* **24**, 979–992 (1998).
57. Auksztulewicz, R., Myers, N. E., Schnupp, J. W. & Nobre, A. C. Rhythmic temporal expectation boosts neural activity by increasing neural gain. *J. Neurosci.* **39**, 9806–9817 (2019).
58. 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. *J. Neurosci.* **33**, 4002–4010 (2013).
59. Rohenkohl, G., Cravo, A. M., Wyart, V. & Nobre, A. C. Temporal expectation improves the quality of sensory information. *J. Neurosci.* **32**, 8424–8428 (2012).
60. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
61. Giordano, A. M., McElree, B. & Carrasco, M. On the automaticity and flexibility of covert attention: a speed–accuracy trade-off analysis. *J. Vis.* **9**, 30.1–10 (2009).
62. Luck, S. J., Hillyard, S. A., Mouloua, M. & Hawkins, H. L. Mechanisms of visual–spatial attention: resource allocation or uncertainty reduction? *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 725–737 (1996).
63. Pestilli, F. & Carrasco, M. Attention enhances contrast sensitivity at cued and impairs it at uncued locations. *Vis. Res.* **45**, 1867–1875 (2005).
64. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
65. Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion decision model: current issues and history. *Trends Cogn. Sci.* **20**, 260–281 (2016).
66. Cheal, M., Lyon, D. R. & Hubbard, D. C. Does attention have different effects on line orientation and line arrangement discrimination? *Q. J. Exp. Psychol. A, Hum. Exp. Psychol.* **43**, 825–857 (1991).

67. Hein, E., Rolke, B. & Ulrich, R. Visual attention and temporal discrimination: differential effects of automatic and voluntary cueing. *Vis. Cogn.* **13**, 29–50 (2006).
68. Ling, S. & Carrasco, M. Sustained and transient covert attention enhance the signal via different contrast response functions. *Vis. Res.* **46**, 1210–1220 (2006).
69. Nakayama, K. & Mackeben, M. Sustained and transient components of focal visual attention. *Vis. Res.* **29**, 1631–1647 (1989).
70. Remington, R. W., Johnston, J. C. & Yantis, S. Involuntary attentional capture by abrupt onsets. *Percept. Psychophys.* **51**, 279–290 (1992).
71. Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nat. Neurosci.* **17**, 347–356 (2014).
72. Reeves, A. & Sperling, G. Attention gating in short-term visual memory. *Psychological Rev.* **93**, 180–206 (1986).
73. Sperling, G. & Weichselgartner, E. Episodic theory of the dynamics of spatial attention. *Psychol. Rev.* **102**, 503–532 (1995).
74. Reeves, A. Attention as a unitary concept. *Vision* **4**, 48 (2020).
75. Bundesen, C. A theory of visual attention. *Psychol. Rev.* **97**, 523–547 (1990).
76. Bundesen, C., Habekost, T. & Kyllingsbæk, S. A neural theory of visual attention: bridging cognition and neurophysiology. *Psychol. Rev.* **112**, 291–328 (2005).
77. Bundesen, C., Vangkilde, S. & Petersen, A. Recent developments in a computational theory of visual attention (TVA). *Vis. Res.* **116**, 210–218 (2015).
78. Jones, M. R. *Time Will Tell: A Theory of Dynamic Attending* (Oxford Univ. Press, 2019).
79. Large, E. W. & Jones, M. R. The dynamics of attending: how people track time-varying events. *Psychol. Rev.* **106**, 119–159 (1999).
80. Vangkilde, S., Coull, J. T. & Bundesen, C. Great expectations: temporal expectation modulates perceptual processing speed. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 1183–1191 (2012).
81. Vangkilde, S., Petersen, A. & Bundesen, C. Temporal expectancy in the context of a theory of visual attention. *Philos. Trans. R. Soc. B* **368**, 20130054 (2013).
82. Anton-Erxleben, K. & Carrasco, M. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nat. Rev. Neurosci.* **14**, 188–200 (2013).
83. Carrasco, M. & Barbot, A. How attention affects spatial resolution. *Cold Spring Harb. Symp. Quant. Biol.* **79**, 149–160 (2015).
84. Lawrence, M. A. & Klein, R. M. Isolating exogenous and endogenous modes of temporal attention. *J. Exp. Psychol. Gen.* **142**, 560–572 (2013).
85. McCormick, C. R., Redden, R. S., Lawrence, M. A. & Klein, R. M. The independence of endogenous and exogenous temporal attention. *Atten. Percept. Psychophys.* **80**, 1885–1891 (2018).
86. Moon, J., Choe, S., Lee, S. & Kwon, O. S. Temporal dynamics of visual attention allocation. *Sci. Rep.* **9**, 3664 (2019).
87. 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. *J. Vis.* **9**, 18.11–14 (2009).
88. Wyart, V., de Gardelle, V., Scholl, J. & Summerfield, C. Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron* **76**, 847–858 (2012).
89. Hilkenmeier, F. & Scharlau, I. Rapid allocation of temporal attention in the attentional blink paradigm. *Eur. J. Cogn. Psychol.* **22**, 1222–1234 (2010).
90. Martens, S. & Johnson, A. Timing attention: cuing target onset interval attenuates the attentional blink. *Mem. Cogn.* **33**, 234–240 (2005).
91. 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. *Atten. Percept. Psychophys.* **76**, 2212–2220 (2014).
92. Di Lollo, V., Kawahara, J.-I., Shahab Ghorashi, S. M. & Enns, J. T. The attentional blink: resource depletion or temporary loss of control? *Psychol. Res.* **69**, 191–200 (2005).
93. Shapiro, K. L., Hanslmayr, S., Enns, J. T. & Lleras, A. Alpha, beta: the rhythm of the attentional blink. *Psychon. Bull. Rev.* **34**, 1472–1478 (2017).
94. 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. *J. Exp. Psychol. Gen.* **134**, 291–307 (2005).
95. Denison, R. N., Parker, J. A. & Carrasco, M. Modeling pupil responses to rapid sequential events. *Behav. Res. Methods* **52**, 1991–2007 (2020).
96. Carrasco, M. in *The Oxford Handbook of Attention* (eds Kastner S. & Nobre A. C.) 183–230 (Oxford Univ. Press, 2014).
97. DeValois, R. L. & DeValois, K. K. *Spatial Vision* (Oxford Univ. Press, 1990).
98. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
99. Kleiner, M., Brainard, D. H. & Pelli, D. G. What's new in Psychtoolbox-3? *Perception* **36**, ECVF Abstract Supplement (2007).
100. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
101. Breitmeyer, B. G. & Ogmen, H. Recent models and findings in visual backward masking: a comparison, review, and update. *Percept. Psychophys.* **62**, 1572–1595 (2000).
102. Acerbi, L. & Ma, W. J. Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Proc. Adv. Neural Inform. Process. Syst.* **30** (2017).
103. Burnham, K. P. & Anderson, D. R. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach* (Springer, 2002).

### Acknowledgements

This research was supported by National Institutes of Health National Eye Institute R01 EY019693 to M.C. and D.J.H., R01 EY027401 to M.C., F32 EY025533 to R.N.D. and T32 EY007136 to NYU. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors thank H.-H. Li for consultation on the model and Carrasco Lab members, especially V. Peña for assistance with data collection and A. Fernández and M. Jigo for their comments on the manuscript.

### Author contributions

R.N.D., M.C. and D.J.H. conceived the project, designed the experiment and interpreted the behavioural results and model findings. R.N.D. and D.J.H. conceived the model. R.N.D. implemented the model, conducted the experiment and analysed the data. R.N.D. wrote and all three authors edited the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41562-021-01129-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41562-021-01129-1>.

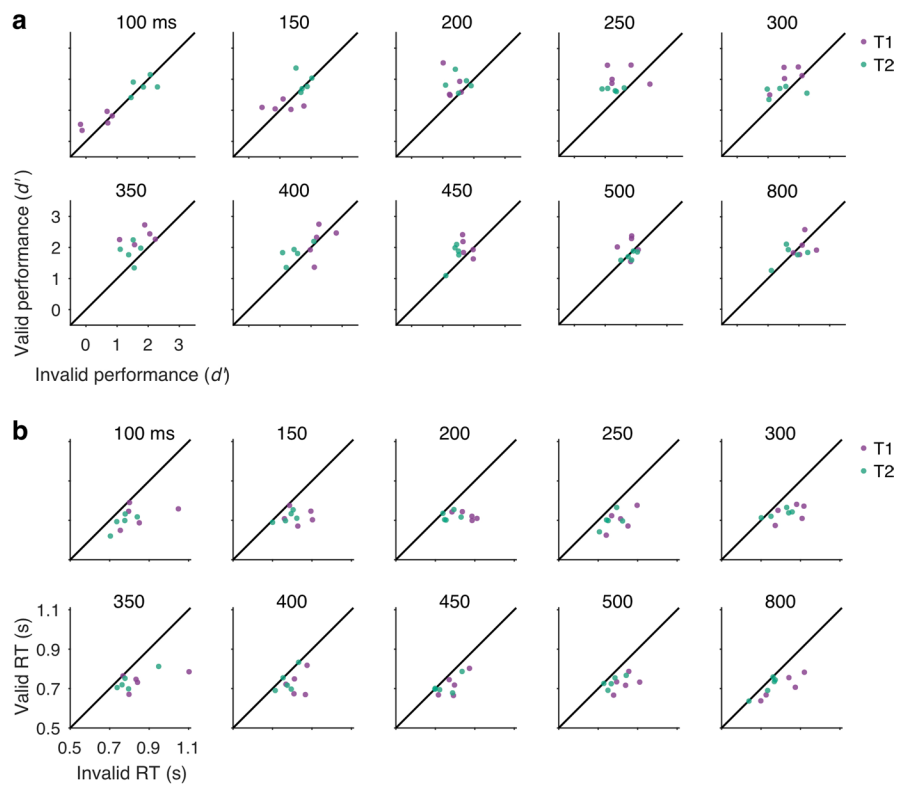
**Correspondence and requests for materials** should be addressed to R.N.D.

**Peer review information** *Nature Human Behaviour* thanks Elkan Akyürek, Benjamin Morillon and Valentin Wyart and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive license to Springer Nature Limited 2021



**Extended Data Fig. 1 | Individual observer data.** Behavioural data for individual observers (data points) at each SOA (separate plots). Valid vs. invalid performance for T1 (purple) and T2 (green). For visualization, individual data across SOAs and cueing conditions were normalized separately for each target by adding a constant to equate the individual target mean with the group mean. This adjusts for individual differences in overall performance for a given target without changing the differences among cueing and SOA conditions, facilitating visualization of the pattern of data across these factors. **(a)** Perceptual sensitivity ( $d'$ ). Data points lying above the unity line have a temporal cueing effect: higher  $d'$  for valid than invalid trials. The improvement of  $d'$  with temporal attention specifically for intermediate SOAs was consistent across individual observers. **(b)** Reaction time (RT). Data points lying below the unity line have a temporal cueing effect: faster RT for valid than invalid trials. Reaction time improvements were consistent across observers.

$d'$ 

Effect	$df_n$	$df_d$	$F$	$p$	$\eta_G^2$
Validity	2	8	8.85	0.0094	0.048
SOA	9	36	3.60	0.0028	0.19
Target	1	4	0.011	0.92	0.00093
Validity x SOA	18	72	0.98	0.49	0.034
Validity x Target	2	8	1.29	0.33	0.010
SOA x Target	9	36	15.38	6.9E-10	0.27
Validity x SOA x Target	18	72	0.99	0.48	0.027

Reaction time

Effect	$df_n$	$df_d$	$F$	$p$	$\eta_G^2$
Validity	2	8	21.92	0.00057	0.27
SOA	9	36	0.41	0.92	0.062
Target	1	4	5.52	0.078	0.091
Validity x SOA	18	72	0.93	0.55	0.018
Validity x Target	2	8	6.83	0.019	0.048
SOA x Target	9	36	1.23	0.31	0.016
Validity x SOA x Target	18	72	0.73	0.77	0.0088

**Extended Data Fig. 2 | Behavioural statistics.** Repeated measures ANOVA table for behavioural data. SOA = stimulus onset asynchrony,  $df_n$  = degrees of freedom in the numerator,  $df_d$  = degrees of freedom in the denominator.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection Data were collected using custom experiment code using MATLAB (version 7.10) with Psychophysics Toolbox (version 3.0). Our experiment code is available on GitHub, <https://github.com/racheldenison/temporal-attention>.

Data analysis Data were analyzed using custom code written in MATLAB (version 9.1). Our modeling code is available on the Open Science Framework (OSF), <https://osf.io/dkx7n>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All behavioral data is publicly available on the Open Science Framework (OSF), <https://osf.io/dkx7n>.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The data are quantitative psychophysical data with a within-subjects, repeated measures experimental design.
Research sample	Five human observers (20-30 years old, 3 female and 2 male) participated in the experiment. 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.
Sampling strategy	We took a psychophysical approach of testing a small number of experienced observers on a large number of repeated-measures conditions (here, 60 conditions). The goal was to model the full psychometric functions, rather than to perform a specific hypothesis test (the typical basis for power analyses). In the vision literature, studies taking this approach typically test 3-5 observers, so we decided to test 5 observers. We initially collected 4.8K trials per observer, based on the number of trials needed to obtain reliable estimates of $d'$ for each condition and observer. For 3 of the observers, we doubled the dataset to investigate a scientific question not considered in the current manuscript, so these observers each had 9.6K trials.
Data collection	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 were generated on an Apple iMac using MATLAB and Psychophysics Toolbox and were displayed on a gamma-corrected Sony Trinitron G520 CRT monitor with a refresh rate of 100 Hz. Observers' heads were stabilized by a chin-and-head rest. Eye position was recorded using an EyeLink 1000 eye tracker (SR Research) with a sampling rate of 1000 Hz.
Timing	We collected data sets from 5 observers with 10 sessions each and 10 additional sessions from 3 of the observers (see "Sampling strategy"). The first session was collected in June 2014 and the last in April 2017.
Data exclusions	No data were excluded.
Non-participation	No participants dropped out or declined to participate.
Randomization	There were no between-subjects experimental groups.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	5 observers, 20-30 years old, 3 female and 2 male
Recruitment	Participants were students or researchers recruited from vision labs at NYU. All were experienced psychophysical observers.



Ethics oversight

University Committee on Activities Involving Human Subjects at New York University

Note that full information on the approval of the study protocol must also be provided in the manuscript.