A common neural network architecture for visual search and working memory

Andrea Bocincova , Christian N. L. Olivers , Mark G. Stokes & Sanjay G. Manohar

Published online: 28 Sep 2020.

Submit your article to this journal ☐

Article views: 81

View related articles ☐

View Crossmark data ☐

A common neural network architecture for visual search and working memory

Andrea Bocincova[a], Christian N. L. Olivers[b], Mark G. Stokes[c] and Sanjay G. Manohar [a]

[a]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK; [b]Department of Experimental and Applied Psychology, Institute for Brain and Behaviour Amsterdam, Vrije Universiteit Amsterdam, Amsterdam, Netherlands; [c]Department of Experimental Psychology, University of Oxford, Oxford, UK

ABSTRACT

Visual search and working memory (WM) are tightly linked cognitive processes. Theories of attentional selection assume that WM plays an important role in top-down guided visual search. However, computational models of visual search do not model WM. Here we show that an existing model of WM can utilize its mechanisms of rapid plasticity and pattern completion to perform visual search. In this model, a search template, like a memory item, is encoded into the network's synaptic weights forming a momentary stable attractor. During search, recurrent activation between the template and visual inputs amplifies the target and suppresses non-matching features via mutual inhibition. While the model cannot outperform models designed specifically for search, it can, "off-the-shelf", account for important characteristics. Notably, it produces search display set-size costs, repetition effects, and multiple-template search effects, qualitatively in line with empirical data. It is also informative that the model fails to produce some important aspects of visual search behaviour, such as suppression of repeated distractors. Also, without additional control structures for top-down guidance, the model lacks the ability to differentiate between encoding and searching for targets. The shared architecture bridges theories of visual search and visual WM, highlighting their common structure and their differences.

The ability to find relevant objects in our visual world relies on maintained knowledge of what to look for. When people are asked to find a particular item, for example with a given colour or shape, the object matching the description must be selected from an array of distracting items. At a neural level, the selection is thought to involve resolving competition between the target and the distractors, by biasing selectivity of feature-sensitive neurons (Bundesen et al., 2005; Desimone & Duncan, 1995; Reynolds & Desimone, 1999). These selection biases are assumed to be established by a mental representation of the item to be searched for, termed an "attentional template". The template information then propagates down the cortical hierarchy through top-down feedback connections, down the cortical hierarchy, pre-activating the template's features (Desimone & Duncan, 1995; Duncan & Humphreys, 1989). This way, stimuli matching the template are selectively amplified for further processing.

A crucial feature of top-down guided search is its flexibility, as a selection is being guided by contexts spanning multiple timescales (Nobre & Stokes, 2019; Woodman & Chun, 2006). Over short time scales, visual working memory (WM) guides visual search behaviour (e.g., Desimone & Duncan, 1995; Pashler & Shiu, 1999). WM allows for the maintenance and manipulation of representations in mind (Baddeley, 1992; D'Esposito & Postle, 2015). To support goal-directed behaviour, it may proactively set up the brain's state to be ready for certain kinds of input-triggered responses (Myers et al., 2017; Olivers et al., 2011), for example, by biasing perceptual selection by automatically amplifying matching perceptual input (e.g., Downing, 2000; Olivers et al., 2006; Soto et al., 2005; Woodman & Luck, 2002). Through these mechanisms, items held in visual WM can act as templates guiding attention. Accordingly, perceptual selection obeys similar laws to a selection of items from WM (Kong & Fournie, 2019; Kuo et al., 2009; Olivers, 2008) suggesting that they could be features of one and the same neural architecture (e.g., Emrich et al., 2009; Mayer et al., 2007).

Indeed in certain situations, visual search and WM appear to overlap. When visual search must be

performed while maintaining a visual object in WM, search is slower when one of the distractors matches the information held in WM (e.g., Moores et al., 2003; Olivers et al., 2006; Pashler & Shiu, 1999). In contrast, this memory-driven interference is absent when multiple objects are held in WM (e.g., Downing & Dodds, 2004; Woodman & Luck, 2007). One way to reconcile these findings is if not all representations in WM are created equal. Specifically, only items that are currently in the focus of attention (FOA) may interact with perceptual information to generate behaviour, and thus drive visual search (Olivers et al., 2011).

While some dedicated models of visual search include working memory as a component (e.g., Bundesen, 1990; Bundesen et al., 2005; Wolfe, 1994), these models do not actually model WM beyond assuming the presence of a biasing representation. Here we take the opposite approach, by starting from the other end and testing a neural model designed exclusively for WM, and assessing how it fares, without modification, on visual search tasks. The correspondence between WM and visual search that emerges from the literature predicts that an architecture built to model WM should in principle generalize to visual search. While taking an off-the-shelf model drastically limits the tasks and effects that can be simulated, it has the benefit of providing a simple, transparent neural mechanism common to both processes.

A recent dynamic model of WM invokes rapid neural plasticity and pattern completion to produce cued recall (Manohar et al., 2019). This model is composed of two kinds of units corresponding to populations of neurons in the brain: a pool of freely-conjunctive neurons – putatively in the prefrontal cortex – rapidly forming associations between the features of an object held in feature-selective neurons (Figure 1A). Even though multiple conjunctive units may initially become responsive to new input, mutual inhibition between conjunction units results in one winning unit eventually becoming selective for the current stimulus. Concurrent activation between the winning conjunction unit and the active feature units drives rapid Hebbian synaptic plasticity, encoding the feature pattern into synaptic connections connecting these units. This produces a stable momentary attractor state which can fall silent when new information is encoded into the network but can be re-activated by a partial cue to recall the object's features via pattern completion. The model's

dynamics only allow one attractor to be in active state at a time; however, the activation state can switched between different attractors corresponding to different memory representations. The attractor landscape, encoded in the network's synaptic weights, is continuously updated: Synaptic weights become more selective for the active pattern of features (corresponding to the item in the FOA), whereas for objects whose attractor has fallen into a silent state (the currently unattended items), the weights are slowly eroded. This model can successfully emulate the capacity limits, decay, primacy and recency effects observed in visual WM.

Can this kind of architecture be harnessed to also implement top-down, template-based visual search? In its current implementation, the activation brought about by sensory stimuli resonates with higher-order areas if the features match an item in memory. In this sense, setting up a template would correspond to encoding information into WM, and visual search would correspond to cued recall in which the test display also contains non-remembered features at non-target locations. To search for a single feature, the template feature is encoded by associating it with the possible locations to be searched, via synaptic plasticity (Figure 1B). During search, recurrent feedback between the prefrontal neurons and the incoming features triggers pattern completion that acts to selectively amplify template-matching features while mutual inhibition suppresses distracting information at non-target locations (Figure 1C). Thus, without any modification, attractor-based working memory models could in principle perform visual search. Indeed other WM models of this class (Fiebig et al., 2020; Fiebig & Lansner, 2017; Mongillo et al., 2008), or the binding-pool model (Swan & Wyble, 2014) which proposes the same general idea, but without the commitment to synaptic plasticity, may also suffice.

To test this possibility, we used the Manohar et al. (2019) model to simulate a set of visual search effects. First, we assessed whether an item in memory (i.e., template) would, via pattern completion, amplify target features and resist distraction from non-target features. Secondly, we examined whether trial history effects stemming from the network's rapid plasticity could simulate benefits related to target repetition effects. Lastly, we asked whether competition between items in memory could account for costs related to the preparation of multiple templates for
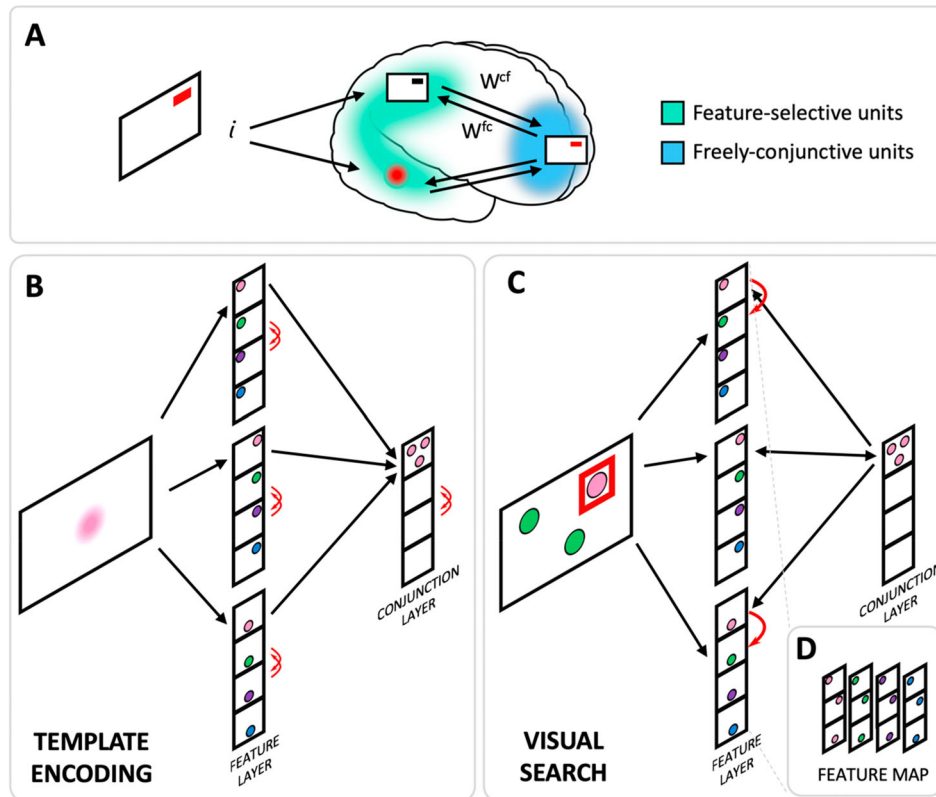
Figure 1. WM model relying on rapid neural plasticity to form object memories can utilize pattern completion and mutual inhibition to implement cued recall and visual search. (A) Depiction of the model's architecture. For the purpose of visual search simulations, the three dimensions within the feature layer were designated to correspond to three searched locations and each unit within a dimension to represent a different colour (as occurs in the primary visual cortex). During memory encoding, the stimulus activates correspondingly tuned units in the feature-selective neurons (posterior cortex). The activation drives activity in the conjunction layer (anterior cortex) where, via mutual inhibition, the conjunctive units compete to encode the combination of active features. The winning conjunctive unit then mutually activates the feature units, and this recurrent activation encodes the information into the synapses via rapid Hebbian plasticity. (B) The same mechanism can be utilized to encode a visual search template into WM. In this case, the association between template colour at all locations is encoded into a conjunction unit. The conjunction unit can be driven by visual inputs to provide recurrent amplification of that colour within the network. (C) The resulting bias within the network guides the activation from the features within the visual search display enhancing the target feature. A successful activation of the template conjunction unit triggers pattern completion that leads to selection of the target and inhibition of the distractors via mutual inhibition. (D) The relative position of feature units within the feature layer can be rearranged to reveal their spatial arrangement within a feature map.

visual search (preparation cost) and costs related to these multiple different templates concurrently guiding the selection of targets (selection cost). Critically, we were interested in the ways in which the model might fail, which would shed light on how WM and search processes differ.

General methods

All simulations used the model's original parameters (Manohar et al., 2019). The data and scripts that support the findings of this study are openly available at https://osf.io/6c9uj. To summarize, the model contains 12 feature-selective units, represented by their activity level (**f**), arranged in three groups of 4, each

connected to 4 freely-conjunctive units (**c**). The activity of each unit is constrained between 0 and 1. Each neuron is self-excitatory and there is a blanket lateral inhibition within each group of units. In its simplest form, the units' activity update equation is as follows is:

$$\begin{bmatrix} c \\ f \end{bmatrix} \leftarrow \text{sigmoid}\left(\begin{bmatrix} \mathbf{W}_{cc} & \mathbf{W}_{cf} \\ \mathbf{W}_{fc} & \mathbf{W}_{ff} \end{bmatrix} \begin{bmatrix} c \\ f \end{bmatrix} + \begin{bmatrix} 0 \\ \text{input} \end{bmatrix} \right)$$

where $\mathbf{W}_{cf}$ and $\mathbf{W}_{fc}$ are conjunctive-feature synapses and $\mathbf{W}_{ff}$ and $\mathbf{W}_{cc}$ are fixed inter-feature and inter-conjunction synapses respectively. The synapses between the feature and conjunctive neurons are continuously updated by a Hebbian covariance rule, to allow

patterns of inputs to be encoded into a conjunctive unit:

$$\Delta = (\mathbf{c} - \beta) \cdot (\mathbf{f}^T - \beta)$$

$$\mathbf{W}^{cf} \leftarrow \sigma(\mathbf{W}^{cf} + \gamma_1 \Delta)$$

$$\mathbf{W}^{fc} \leftarrow \sigma(\mathbf{W}^{fc} + \gamma_2 \Delta^T)$$

where Δ is the change in firing rate, β is the units' baseline activity, γ are the learning rates and σ is a function that constrains values to lie between 0 and 1.

The original feature arrangement was designed to showcase a general feature-binding mechanism and does not treat space as special. In order to model visual search, the feature-selective units were relabelled to represent a spatial feature map, with four colours appearing at three possible locations (Figure 1D). Spatial information is integral to visual search, so here we incorporate it into colour features in the form of a map. In fact, it is impossible to present a search array without spatially-specific neurons, and accordingly all previous models of visual search have used spatial feature maps. This exactly preserves the structure of the original plastic attractor network. To encode a template into the network, the target colour was presented simultaneously at all possible locations, creating an attractor state for that colour via a flexible conjunction unit. This unit could then function like a top-down rule, signalling the need to search for a specific target colour.

Simulation 1: Distractor effects

The efficiency of visual search is limited by the amount of competition between the target and distractors. First, this competition is typically expressed as a set size effect, i.e., the time necessary to locate the target as a function of the number of distractors present (Wolfe, 1994, 1998). Second, competition varies as a function of target-distractor similarity, as well as distractor-distractor similarity (e.g., Bauer et al., 1996; Duncan & Humphreys, 1989; Humphreys et al., 1989; Pashler, 1987; Wolfe et al., 1992). Third, distractor competition has been shown to be reduced by repetitions in distractor locations and identity, suggesting that search is aided by distractor suppression (Hout & Goldinger, 2010; Kristjánsson & Driver, 2008).

The primary goal of the first simulation was to assess whether the model can perform visual search, i.e., select targets among distractors. In addition, we evaluated how changes in distractor-related factors affect the model's ability to select targets. We first manipulated the *number* of distractors presented with target (0, 1 and 2 distractors) to assess any basic set size effects. Second, we tested whether distractor-distractor similarity (i.e., a homogeneous versus heterogeneous distractor set) increases search times. Finally, we assessed how the repetition of distractor location and colour from trial to trial reduced search times.

Method

Simulation

To simulate the visual search set size effect, the target colour was presented with zero, one or two distractors. The simulated trial sequences for these three different experimental conditions are shown in Figure 2A. Each trial started with a pre-stimulus period lasting 200 time steps [ts] with all features under maximal inhibition (input $i = -1$). During the following encoding epoch, the template was encoded by activating the corresponding feature units, by maximally activating the features corresponding to the template search-defining feature (i.e., colour) and its possible locations ($i = +1$) while suppressing all the other features ($i = -1$). A delay period ($i = 0$) of 300 ts followed template encoding. The visual search display (120 ts) was simulated by activating the template search-defining feature unit at a target location ($i = +1$, $i = -1$ for all other feature units). The search display was followed by a short period of feature layer inhibition (5 ts) and a response period (400 ts). The inhibition period served to temporarily suppress the activation related to the visual search display and allow the dynamics of the network to generate a response. Without this suppression, the responses would be driven by the external input rather than the workings of the model. To simulate the set size effect, the target was presented either on its own or together with one or two other colours presented at the remaining locations. The number of distractors varied between 0 and 2 on a trial-to-trial basis. The model was exposed to the trial sequences in blocks with each block consisting of 300 randomized trials (100 trials for each distractor number). A total of 100 simulations were performed.
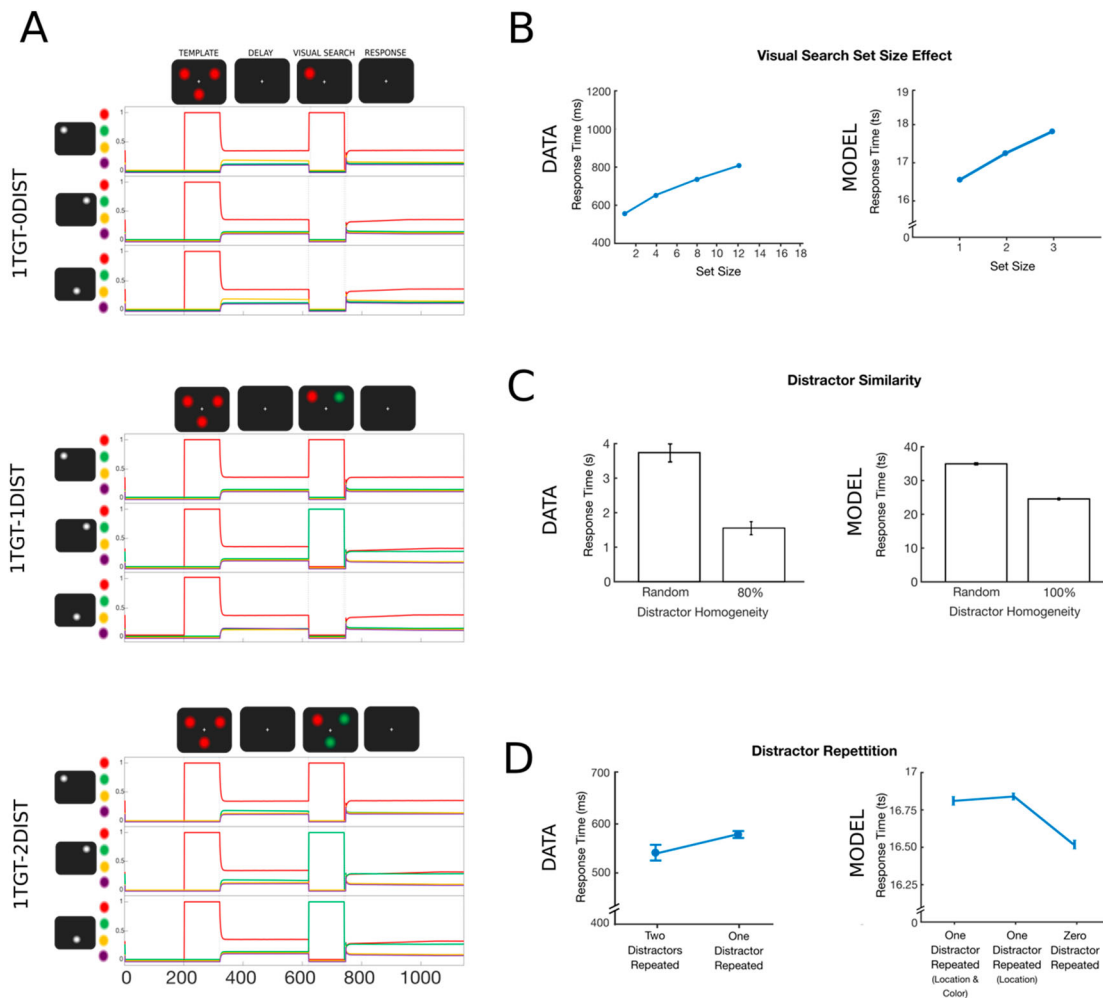
Figure 2. Simulation 1: Visual search set size effect, distractor repetition and distractor similarity. (A) Time course of feature activity traces of a representative trial for the three experimental conditions. Each trial started with a 200 ts inhibition period followed by encoding of the template (120 ts). Following a delay period of 300 ts, visual search display was presented for 120 ts. Responses were recorded over the following response period of 400 ts. A variable number of distractor colours were presented at the two non-target locations (zero- top, one- middle, two- bottom). For visualization, a small vertical shift was added to the feature neuron firing rate traces. (B) The model reproduced the visual search set size effect. Increasing the number of items presented during visual search led to an increase in RTs (set size 1 = zero distractors). Data were reproduced from Wolfe (1994). (C) Search was more efficient for homogeneous compared to heterogeneous distractors. Data were reproduced from Kong et al. (2016). (D) The model, however, did not replicate the distractor repetition effect. In the 1TGT-1DIST condition, repetition in the distractor's location or its location and colour in addition to the target location did not improve RTs compared to trials in which the distractor changed its location (i.e., zero distractor repeated). Data were reproduced from Kristjánsson and Driver (2008).

In the above simulation, there is no hard-wired relation between the same colour at different locations, and so the network cannot recognize distractor colour similarity. To train the model to recognize colour similarity, in a separate simulation the model was exposed to 100 trials during which, for each colour, the units at all three locations were associated with each other (e.g., green-left, green-right and green-bottom were co-activated). Following the training, the model was exposed to a mixture of two types of trials: trials with either homogeneous

distractors (e.g., green at two locations) or heterogenous distractors (e.g., green and purple).

Behavioural performance from simulated data

Performance was measured in two ways: accuracy and response time. The model's response was considered correct if the most active unit within the target location (i.e., dimension) corresponded to the template's colour. Response time (RT) was established as the time point during the response period at which the most active unit reached its maximum level of

activation. Qualitatively similar results were obtained when all feature units were considered, i.e., if a response was considered correct when the most active unit out of *all* 12 feature units (or two most active units of all 12 features for two-target conditions) corresponded to the colour *and* location of the target.

Distractor repetition effect was only evaluated with a set size of 2 items (1TGT-1DIST). Because the model architecture only includes three locations, if all were occupied in the display as was the case in set size of 3 items (1TGT-2DIST), it would not be possible to evaluate the benefit of a single distractor repetition independently of the cost associated with target location switch (i.e., any change in the location of a distractor in this condition would produce a change in the target location). If locations of both distractors were repeated, the target location would necessarily have to be repeated as well, making it impossible to compare this condition to the single distractor repeat condition.

Results

As shown in Figure 2B, the model produced a set size effect, showing an increase in RTs as more distractors were presented during visual search. The zero-distractor condition (set size one) represents the time necessary for the model to perform a cued recall. In line with empirical findings, the more distracting features that were added to the search display, the more time was necessary for the target to become active. Because accuracy was at ceiling, no accuracy differences were observed between conditions. After the model was trained to recognize distractor colour similarity, the model produced longer RTs for trials with heterogeneous distractors compared to homogeneous distractors (Figure 2C).

The model did not reproduce the distractor repetition effect. In contrast to human data demonstrating faster search with more repeated distractors (Kristjánsson & Driver, 2008), the model was slower to search when there was repetition of a distractor in addition to the target (Figure 2D).

Interim discussion

The goal of the first simulation was to establish that the WM model can in principle perform visual search. We assessed whether pattern completion –

the tendency to re-activate recently-encountered patterns of input – could be utilized to select targets and overcome distraction during visual search. To do so, we explored how well the model matches human performance by manipulating distractor-related factors (the number of distractors, distractor repetition, distractor similarity). In line with the empirical data, the model was worse at selecting a target when more distractors were present (Wolfe, 1994, 1998). This was because each distractor decreased the selectivity of the signal produced by the target unit activation by indiscriminately activating non-template conjunction units. This competition led to a delay in the recurrent activation between the target and the template.

Additionally, the model was worse at rejecting distractors when the distractor was repeated from one trial to another, failing to produce distractor repetition benefit (Hout & Goldinger, 2010; Kristjánsson & Driver, 2008). This is because distractor information was erroneously encoded into the template's conjunction unit when concurrently presented with the target during trial *n* search making the competition for the target unit stronger on trial *n+1*. Since the model does not have any gating or executive control mechanisms, it does not distinguish between phases of the tasks, and treats encoding and search identically. As such, the network cannot prevent distractors from becoming integrated into the template representation following a search.

The failure of the model, where distractors erroneously get associated with the template, highlights that true visual search may involve an additional mechanism not present in WM. Non-matching information may be suppressed, perhaps through heightened lateral inhibition between features, driven by contextual knowledge about the task.

It could be argued that, because the model's current architecture does not allow true conjunction search, which is typically associated with set size effects (Wolfe, 1994), a visual search slope should not be observed in our simulation. However, simple feature search is less efficient when the similarity between distractors decreases (Bauer et al., 1996; Humphreys et al., 1989; Wolfe et al., 1992). Because the model does not have any a priori colour knowledge, e.g., it cannot know when the blue-left unit matched the blue-right unit, it naturally performs a visual search involving heterogeneous distractors. To

test whether the model is more efficient at selecting a target among homogeneous distractors, we associated feature units coding for the same colour across the three locations which allowed the model to recognize when the same colour distractor was presented at two different locations. Following this training, the model was more efficient at selecting a target among homogeneous compared to heterogeneous distractors. This may explain why the simple feature search produced a visual search set size effect.

Encoding colour associations into the network's synapses to model distractor similarity led to slower search. After colour associations were created, distractors could activate a conjunction unit corresponding to their colour in the same way as the target colour activated the template conjunction unit, which resulted in stronger competition with the template colour. This apparent idiosyncrasy of the model may actually correspond to empirical findings suggesting that it is very difficult to use a template to ignore rather than facilitate matching input (Noonan et al., 2016, 2018).

Simulation 2: Target repetition effect

Simulation 1 showed that distractor repetition allows the distractors to establish biases within the network, which are detrimental to performance. However, repetition of *target* identity has been shown to aid search (target repetition effect; Hillstrom, 2000; Maljkovic & Nakayama, 1994). The model is particularly well adapted to produce task-relevant repetition benefits due to its rapidly plastic synapses. The longer a template is represented by an active conjunction unit, the stronger it is encoded. Because the strength of the synapses evolves over time, history influences the behaviour of the network across trials. Simulation 2 aimed to test whether target repetition aids performance.

Method

Simulation

To simulate target repetition effects, two templates were encoded sequentially (encoding period of 120 ts, followed by an inter-stimulus interval of 500 ts with no input $i = 0$) followed by two visual searches (Figure 3A). The targets used in the two searches were either of the same colour (repeat trial) or the

templates switched colour from trial one to trial two (switch trial). Each simulation consisted of 200 randomized trials for each condition. A total of 100 simulations were performed.

Result

In line with empirical findings, repetition in the template colour in two consecutive visual searches led to lower error rates and faster RTs compared to when the template colour was switched between trials (Figure 3B). Repetition of the location of the target in addition to its colour improved accuracy of target selection, as expected, however, it *prolonged* RT (Figure 3C). When the target colour changed on the next trial, then location repetition had the opposite effect: it produced faster, more erroneous responses.

Interim discussion

Simulation 2 successfully replicated the target repetition effect showing fewer errors and faster responses when a template was repeated across two searches. The benefit of target repetition was driven by repetition of the template's colour (Hillstrom, 2000; Maljkovic & Nakayama, 1994), because the repetition strengthened the amplification produced by the conjunction unit across all searched locations. Repeating the target location did *not* consistently facilitate search on same-target trials, in contrast to location priming findings (Maljkovic & Nakayama, 1996; Rabbitt et al., 1979; Treisman, 1992). After the first search, the selected template is more strongly associated with its colour at the target location. But also, the *non*-selected template colour is more strongly associated with its colour at that location because of mutual inhibition in both feature and conjunction layers. This double effect results in increased competition from the other template colour, but also a bias to report the same colour, for that same location. At the other locations, however, the associations are not affected as much by the first search. Together, this generates the counterintuitive pattern of errors and reaction times.

The failure to reproduce location priming suggests that visual search may also include an additional *feature-general* spatial map – for instance a global salience map – which is missing from our model.
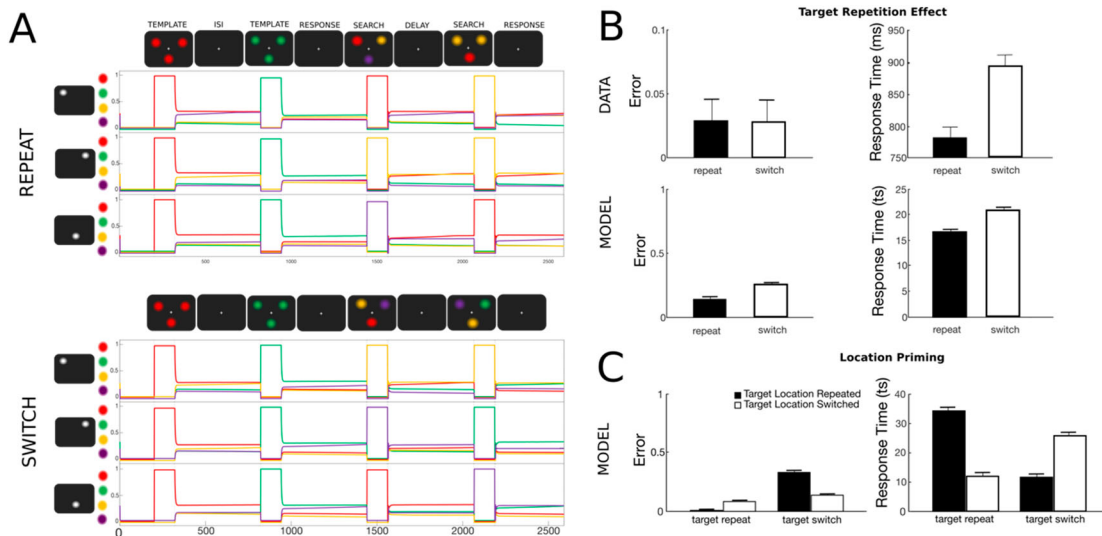
Figure 3. Simulation 2: Visual search target repetition effect. (A) Time course of feature activity traces of a representative trial for the two experimental conditions. Each trial started with a 200 ts inhibition period followed by the encoding of the first template (120 ts). Following a delay period of 300 ts, the first visual search display was presented for 120 ts. This sequence was repeated for the second template (template repeat- top, template switch- bottom). Responses to the second visual search were recorded over the response period of 400 ts. Two distractor colours were always presented at the two non-target locations. For visualization a small vertical separation has been added to the feature firing rate traces. (B) Target repetition leads to improved performance compared to template switch. Both the number of errors (top) and RTs (bottom) were higher on trial pairs in which the template switched its colour compared to trials pairs in which the template stayed the same. Data were reproduced from Maljkovic and Nakayama (1994). (C) Location priming in target repetition and target switch trials. Repetition of target location in addition to its colour resulted in higher accuracy, but target selection was delayed compared to when the target was presented at a different location. In contrast, when target colour was switched from one trial to another, the model produced more errors, however when the correct target colour was selected, it was selected faster.

Simulation 3: Differences in costs related to preparing multiple templates and using multiple templates to select targets during search

A WM model may be particularly apt for simulating the capacity limits related to the number of items that can currently serve as an attentional template, i.e., the number of targets that one can search for. WM capacity is limited to a few (nominally 3–5) objects (Bays, 2015; Cowan, 2001; Luck & Vogel, 1997; Oberauer et al., 2016), but only a subset of these items is prioritized for action (Cowan, 1999; Cowan et al., 2005; Oberauer, 2002; Souza & Oberauer, 2016). Indeed, most WM tasks require only one item to be activated at once, at the time of response.

Similarly, in visual search literature, this same question arises with respect to the number of templates that can concurrently guide target selection. Prior visual search studies report conflicting results, sometimes favouring a strict limit of a single item template (Olivers et al., 2011; van Moorselaar et al., 2014) or at other times providing support for attentional selection

driven by multiple templates (Bahle et al., 2018; Hollingworth & Beck, 2016; Zhang et al., 2018).

One possible resolution of this dilemma could be that different capacity bottlenecks act on different stages of visual search. According to a study by Ort and colleagues (2019), the limit imposed on the number of items that can be *concurrently prepared* for visual search (i.e., prior to display onset) may be different from the limit imposed on the number of templates that can be used to *concurrently select* targets (i.e., extracting those targets from the actual display). In other words, they proposed a distinction between the question of multiple template preparation and multiple template-based selections. In this study, participants remembered either one or two templates (1TMP vs 2TMP) and searched for two targets which either both matched one of the templates (1TGT, so both targets were the same), or each matched one of the two templates (2TGT, so the targets differed). The results showed only a small decrement in behavioural performance resulting from maintaining an additional search template

(multiple template preparation cost: 1TMP-1TGT vs 2TMP-1TGT). Crucially, behavioural performance suffered considerably more when the two templates both had to be matched to the search targets, compared to when only one of the two remembered templates had to be used to find the targets (multiple target selection cost: 2TMP-2TGT vs 2TMP-1TGT). Based on these results, the authors argued that the crucial capacity limitation during visual search arises due to a selection bottleneck during template-guided prioritization (i.e., the selection stage) rather than due to limitations imposed on the number of items serving as attentional templates (i.e., the template activation stage). In their account of the neural underpinnings of these capacity limitations, they hypothesized that even if two coexisting templates successfully establish two different bias signals, the concurrent activation of the templates is slowed down via mutual inhibition when the input contains features from two different templates.

The goal of Simulation 3 was to test whether competition between the model's units can account for the differences in cost related to the number of prepared templates (multi-item template) and the number of templates used concurrently to select targets (multi-template search). The first question we wanted to address in Simulation 3 was whether there is a cost to preparing multiple search templates and whether attended as well as unattended templates can guide attention. To allow separation of object representations within the model (i.e., encoding of a single object into a single conjunction unit), the model requires information to be encoded into memory sequentially. As a consequence of serial encoding, the conjunction unit for the first encoded template falls silent to allow the second template to be encoded. Even though the unit falls silent, the biases established by the first template persist in the network's synapses. Thus, even though the two templates are not co-active, they are both still able to guide selection during visual search, but at different levels of preparation. The cost in performance caused by increasing the number of templates emerges from competition between the conjunction units encoding the templates, and the effect this competition has on their bias signals during their retention. The second question we wanted to address was whether there is a cost to target selection driven by multiple templates compared to a single template and whether this is achieved through a parallel activation of the templates.

Method

Simulation

Trial sequences were simulated as described in Simulation 2. A period of inhibition (50 ts) was added prior to the visual search (Figure 4A). This served to bring down the activation level of the second encoded template and allow the first template to become activated by the features in the visual search display.

The simulation included four possible experimental conditions. We varied the number of templates (1 or 2 TMP, multiple template cost) that were encoded into the network and the number of templates used to select the two targets (1 or 2 TGT, multiple template-based selection cost). For condition 1TMP-1TGT, the template was encoded twice in a row to temporally match the other two conditions containing two templates, 2TMP-1TGT and 2TMP-2TGT. For the 2TMP-1TGT condition, 50% of trials probed the first encoded item, 50% of trials probed the second encoded item. The model was exposed to the trial sequences in blocks with each block consisting of 100 trials of one condition type (2 blocks per condition). The same template(s) was used across all trials within a block as in the original design (Ort et al., 2019). A total of 100 simulations was performed.

Behavioural performance from simulated data

A trial was considered correct if the most active units at the two target locations corresponded to the template colour and RTs were measured as the time point at which both units reached their maximum level of activation. RTs were calculated for correct trials only.

Results

In line with the empirical findings, the simulated data shown in Figure 4B provided evidence for different costs associated with setting up two versus one attentional template (multiple template preparation: 1TMP-1TGT vs 2TMP-1TGT) and selecting targets based on one or two templates (multiple template-based selection: 2TMP-1TGT vs 2TMP-2TGT). In particular, the results showed a relatively small cost associated with preparing two versus one template for search
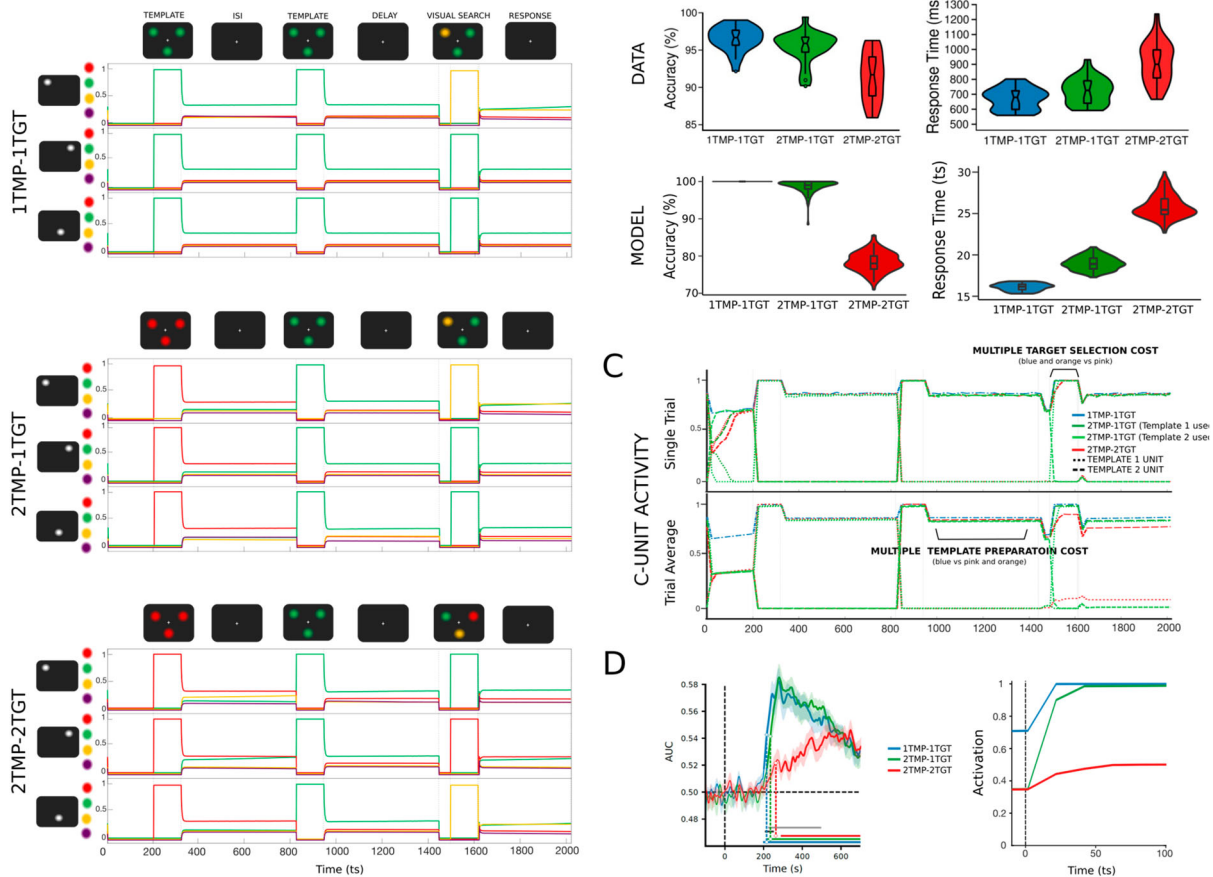
Figure 4. Simulation 3: Multiple template preparation cost and multiple template-based selection cost. (A) Time course of feature activity traces of a representative trial for the three experimental conditions. Each trial started with a 200 ts inhibition period followed by a serial encoding of one or two templates (120 ts; ISI 500 ts). Following a delay period of 300 ts, all features were inhibited (50 ts) before and after (5 ts) visual search display was presented for 120 ts. A distractor colour was always presented at a non-target location. For visualization a small vertical separation has been added to the feature unit activity traces. (B) Multiple template preparation cost: the activation of a second template in preparation for search resulted in reduced accuracy: 1TMP-1TGT (blue) vs 2TMP-1TGT (green). Accuracy was worsened even more when both templates had to be engaged during search: 2TMP-1TGT (green) vs 2TMP-2TGT (red). Decrements in accuracy were accompanied by increases in RTs. The violin plots represent the distribution of accuracy (A) and RTs (B) derived from empirical data (top) and simulated data (bottom). The boxes represent the median (central horizontal bar) and quartile range (upper and lower horizontal bar). The vertical lines represent the minimum and maximum (lower and upper quartile + 1.5 * interquartile range, respectively). Outliers are depicted as single unfilled dots. (C) Single unit and average conjunction unit activity representing the templates for the different experimental conditions. *Multiple template preparation cost:* During preparation for visual search, there was a small cost associated with preparing two templates each encoded in a separate unit compared to only preparing one template (blue vs green). When currently active template had to be switched to perform search, as was the case when the target matched the first encoded template, there was a slight delay in the template unit's activation during search. *Multiple target selection cost:* There was a more significant cost associated with an increased number of templates (and conjunction units) engaged to guide visual search. In the 2TMP-2TGT condition (red), the search-guiding conjunction unit took longer to reach its peak activation due to competition with other conjunction units. The average trace suggests concurrent and weaker activation of both templates (red dashed and dotted lines). However, the examination of single unit traces showed activation of only one of the two conjunction units (different unit on different trials). (D) Similarly to empirically recorded neural data, average activation traces of template conjunction units showed a smaller template preparation cost (1TMP-1TGT vs 2TMP-1TGT). However, the average activation of the templates' neural representations in the 2TMP-2TGT condition was significantly delayed corresponding to the added cost of target selection based on two different templates.

(2TMP-1TGT vs 1TMP-1TGT) in both accuracy and RT. Average template conjunction unit activation during template preparation showed less activation in the two-template condition compared to the one-template condition (Figure 4C). Additionally, the empirical

data of Ort et al. (2019) also showed that the cost associated with preparing two templates was stronger when the target matching template changed from one trial to another, compared to repetitions. This was replicated in the simulated data, as repetition in

the colour of the target increased the stability of the template's stable attractor state, improving accuracy and RT compared to switch trials (see also the previous simulation for repetition benefits).

Crucially, an even larger decrement in performance was observed as a result of multiple template competition during target selection, i.e., when the number of templates matching targets increased from one to two: 2TMP-1TGT versus 2TMP-2TGT. This is again in line with the empirical results of Ort et al. (2019) and suggests that simultaneously selecting the search targets based on two different templates slows down RTs and impairs the accuracy with which the targets are selected.

Interim discussion

The goal of Simulation 3 was to test whether mutual inhibition implemented in the model to separate object representations could also produce competition between multiple templates during preparation for search as well as during search itself (Ort et al., 2019). Simulation showed a small cost to accuracy and RTs when maintaining two versus one template. This cost arose from competition between the conjunction units corresponding to the two templates. It can be likened to the set size effect observed in WM tasks when performance decreases with the number of objects maintained in WM. Competition between templates arises at the time of template encoding and maintenance, resulting in the multiple template preparation cost.

The current model has an FOA limited to a single item. The fact that it can replicate a visual search task that requires the preparation of multiple templates means that, in the model, unattended items can still also guide attentional selection. Ort and colleagues (2019) reported improved performance for 2TMP-1TGT trials when the template used to select the target was repeated between two trials. It could be the case that the most recently used template remains in the FOA and so is more efficient at guiding visual search. Such intertrial effects were also present in the simulated data and were related to the amount of time a particular template spent in the FOA. Even though the template might not be in the FOA at the time of visual search, if this template was used on the previous trial to select the target and so was attended for longer on a previous trial, the strengthened bias signal creates an implicit guidance in addition to the explicit guidance driven by the current FOA. So, one

might say that the bottleneck in search template activation is due to one and the same mechanism as the bottleneck of prioritizing a single item in WM.

In line with the empirical data, even when the number of active templates was kept constant (equal multiple template preparation cost), actually searching for two targets matching different templates was costly. This multiple template search cost was driven by competition during pattern completion when input engaging both templates at the same time led to slower activation of the conjunction unit (Figure 4 (C)). An initial examination of the average conjunction unit activity on correct trials suggested that both templates were activated concurrently during visual search. However, single-trial activity showed that only one of the template conjunction units remained active after overcoming the initial competition. But how, then, can the model select two different colour targets simultaneously? To address how the model succeeded in this, we examined the synaptic weights. Unlike in any other condition, the conjunction unit representing the first template had stronger synaptic connections with targets associated with the second template, and vice versa. This allowed the model to correctly select two different coloured targets at two different locations even if only one conjunction unit was active. Such shared selectivity between the two conjunction units representing templates suggests that simultaneously driving both templates caused them to partially merge into a joint template. Indeed, it is possible that, due to the blocked nature of the experimental design, participants in the experiment also formed a partially overlapping representation of the two template colours when both colours were needed to select targets.

General discussion

The goal of the current study was to test whether an existing model of WM can be generalized to model behaviours that it was not designed to perform. Specifically, we tested whether a single architecture could support WM and visual search, two processes that carry a significant overlap in their mechanisms. Despite their commonalities, these processes differ. Here we addressed a gap in the current literature by modelling both within one architecture, and we find that this difference is non-trivial.

We show that an existing neural model, relying on rapid plasticity to form WM representations and pattern completion to produce cued recall, was able to employ these same mechanisms to execute visual search. In this model, a search template is encoded into the network as synaptic weights, connecting a flexibly-conjunctive unit with feature units representing the template feature at searched locations. These weights generate a bias that guides recurrent activation during visual search, amplifying feature units containing the target while suppressing distractors. Without modification, still using parameters fitted to match WM data, the model reproduces patterns of results in visual search data it was never designed to model. Particularly, the model produced the set size effect (Wolfe, 1994), target repetition effect (Hillstrom, 2000; Maljkovic & Nakayama, 1994) and showed how mutual inhibition, originally implemented to separate items in WM, can also operate between templates to resolve competition during visual search (Ort et al., 2019). In addition to simulating behaviour, the model can also provide insights and predictions about the neural activity behind it.

One key difference between WM and search is the direction of bias at the time of test. Visual search involves a bias generated by the search template that allows participants to find task-relevant features in the display, whereas in WM, a test cue biases the re-activation of associated WM representations. Our model's bidirectional interactions allow both effects, but also make predictions about situations when there are multiple search templates. Traditional accounts of search emphasize top-down guidance and bottom-up salience but do not incorporate mechanistic accounts of selecting between multiple templates. Our model allows for maintained representations to exist in different levels of prioritization. Thus, a search array can help select one of the templates by bringing it into FOA, much like in WM recall, while the templates simultaneously guide search through their synaptic biases even if currently in a passive state.

Although the model was able to simulate crucial aspects of visual search, it fails to capture other characteristics. The interpretations of these failures can be grouped into two types: some reflect mechanisms common to WM and search that could be implemented in future; whereas others might reflect true differences between WM and search.

Failures reflecting unimplemented common aspects

One problem is that the model does not implement feature-based attention. To simulate visual search, each target-defining feature first had to be encoded together with the response-defining features (e.g., the possible locations at which this feature could appear during search). Additionally, due to its small number of units, the model's current architecture can only simulate single-feature templates and three possible target locations, which limits the model's ability to perform certain types of visual search (e.g., classic conjunction search). The current arrangement of features also does not allow the model to recognize colour similarity between different locations (i.e., the presence of the same colour at two locations). A parallel limitation for WM would be that the spatial arrangement of features adopted here would abolish misbinding of spatial features. Including long-term plasticity or non-spatial feature neurons could fulfil this role. It could be the case that in V1, colours are coded in location-specific colour-selective neurons, whereas in higher visual areas there are location-general colour-selective neurons that are activated by a colour *wherever* it appears.

Furthermore, the current WM model ignores low-level bottom-up visual processing that is of great importance in visual search. Pop-out and singleton detection presumably require additional visual domain-specific circuits, which are ignored in our model but could also influence WM encoding. The omission reflects a tendency of some WM research to focus on high level binding and retrieval mechanisms, of a domain-general kind.

Failures reflecting underlying differences between WM and search

The second set of limitations is informative regarding the differences between WM and search. For example, the model was unable to take advantage of repeated distractors to speed up the detection of the target. Previous literature suggests that observers store contextual information (e.g., distractors and their location) and that this information affects future search (Chun & Jiang, 1998; Geyer et al., 2006; Hout & Goldinger, 2010). Despite the ability of the model to store information about prior distractors, the model was

unable to use this information in an effective way, i.e., the model was not able to reject distractors if they were repeated. This is because the distractors were encoded into the template representation, which caused them to become facilitated during search. Note that empirical findings suggest that it is in fact very difficult to use a template to ignore rather than facilitate matching input (Noonan et al., 2016, 2018).

A second distinction highlighted by the model is that visual search may entail a degree of understanding of the task structure, to unambiguously distinguish cue encoding (setting up the template) from search (applying the template). The model has no such distinctions, and the simple attractor dynamics are consistent throughout the task. In working memory recall, a single feature is activated to probe the network, drawn from one of the previously encoded objects. This engages a conjunction unit and, through pattern completion, activates one of the network's stable attractor states. However, to simulate visual search, instead of providing the model with a single cue to trigger recall, the memory test (i.e., the visual search display) consists of a *new combination* of features, corresponding to both template and distractors. Similarly to a recall cue, the target feature tries to re-activate one of the stable states, corresponding to the templates, opening up a possibility that visual search could also be framed in terms of pattern completion. However, because the model lacks a gating mechanism, it currently cannot differentiate between the need to engage in visual search and the need to encode a template. As a result, the novel combination of features in the visual search display sometimes becomes encoded into a new conjunction unit instead of triggering pattern completion. We conclude that human visual search may employ an *additional gating mechanism* that is unnecessary for basic WM tasks. Our results highlight that simple lateral inhibition alone would not account for distractor gating, although gating mechanisms proposed for WM could potentially achieve this (Frank et al., 2001; O'Reilly & Frank, 2006).

Due to these inherent limitations, the WM model cannot outperform dedicated models of search. The value of this study lies in demonstrating that a WM model can generalize to a task it was not designed to perform. Furthermore, its failures critically help to demarcate where the processes of visual search differ from those of memory recall. There are

Table 1. Brief descriptions of a selection of prior visual search findings and the ability of the model to simulate them.

Effect	Data	Model
Search Efficiency	Steeper RT slope for inefficient visual search	☑ Slower reaction times for inefficient search (visual search with heterogenous distractors – Simulation 1)
Distractor Repetition	Faster search RT for consecutive trials containing distractors at the same locations	☒ Repetition of distractor led to RT slowing due to distractor encoding (Simulation 1)
Bottom-Up Search	Unique (singleton) or salient (pop-out) items are detected faster	☒ Model lacks low level sensory processing and cannot account for low-level visual factors in visual search
Top-Down Guidance	WM template guided search and attentional capture	⑦ The model produces facilitation of features matching an item in WM (Simulation 1–3). Despite not being directly tested, items maintained in WM should automatically capture attention (i.e., elicit a shift of FOA)
	LTM representations influence visual search	⑦ The model currently only relies on short-term synaptic plasticity, however, training the model to represent colour similarity through repeated exposure prior to simulation showed the potential of the model to incorporate longer-term memory components (Simulation 1). Adding LTM plasticity could account for LTM-based effects in visual search
Visual Search Capacity Limitations	There is a limit to the number of items that can serve as an attention template	☑ Mutual inhibition within the model can account for limitations related to the number of items that can concurrently serve as an attentional template as well as for the limits in the number of targets that can be concurrently selected (Simulation 3)
Visual Search History	Repetition of target-defining feature speeds up target detection	☑ Repetition of target colour facilitated target detection (Simulation 2)
		☒ The model currently cannot account for target location priming (Simulation 2). This could suggest that visual search additionally recruits a feature-general salience map

Notes: Tick marks correspond to successfully replicated effects, cross marks correspond to effects that the model currently cannot simulate or failed to replicate, and question marks correspond to untested effects that the model is likely to reproduce.

additional visual search effects that the model cannot currently account for (see Table 1), but expanding this model to simulate other currently missing

aspects of WM might allow it to model visual search better as well. Bringing WM and visual search under one common architecture is an important step in modelling flexible behaviour. Future efforts should be focused on improving the ability of cognitive models to incorporate multiple cognitive processes that share common mechanisms within one framework. During this integration, it will, however, be important to consider the true overlap between the functions in question, and separately model components that can be accounted for by other mechanisms.

Disclosure statement

Funding

ORCID

Sanjay G. Manohar 🄳 http://orcid.org/0000-0003-0735-4349

References

Baddeley, A. D. (1992). Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience*, *4*(3), 281–288. https://doi.org/10.1162/jocn.1992.4.3.281

Bahle, B., Beck, V. M., & Hollingworth, A. (2018). The architecture of interaction between visual working memory and visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(7), 992–1011. https://doi.org/10.1037/xhp0000509

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly separable from distractors. *Vision Research*, *36*(10), 1439–1466. https://doi.org/10.1016/0042-6989(95)00207-3

Bays, P. M. (2015). Spikes not slots: Noise in neural populations limits working memory. *Trends in Cognitive Sciences*, *19*(8), 431–438. https://doi.org/10.1016/j.tics.2015.06.004

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*(4), 523–547. https://doi.org/10.1037/0033-295x.97.4.523

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, *112*(2), 291–328. https://doi.org/10.1037/0033-295X.112.2.291

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185. https://doi.org/10.1017/S0140525X01003922

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. https://doi.org/10.1016/j.cogpsych.2004.12.001

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Downing, P. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473. https://doi.org/10.1111/1467-9280.00290

Downing, P., & Dodds, C. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*(6), 689–703. https://doi.org/10.1080/13506280344000446

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Emrich, S. M., Al-Aidroos, N., Pratt, J., & Ferber, S. (2009). Visual search elicits the electrophysiological marker of visual working memory. *PLoS One*, *4*(11), e8042. https://doi.org/10.1371/journal.pone.0008042

Fiebig, F., Herman, P., & Lansner, A. (2020). An indexing theory for working memory based on fast Hebbian plasticity. *eNeuro*, *7*(2). https://doi.org/10.1523/ENEURO.0374-19.2020

Fiebig, F., & Lansner, A. (2017). A spiking working memory model based on hebbian short-term potentiation. *Journal of Neuroscience*, *37*(1), 83–96. https://doi.org/10.1523/JNEUROSCI.1989-16.2016

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160. https://doi.org/10.3758/CABN.1.2.137

Geyer, T., Müller, H. J., & Krummenacher, J. (2006). Cross-trial priming in visual search for singleton conjunction targets: Role of repeated target and distractor features. *Perception & Psychophysics*, *68*(5), 736–749. https://doi.org/10.3758/BF03193697

Hillstrom, A. (2000). Repetition effects in visual search. *Perception and Psychophysics*, *62*(4), 800–817. https://doi.org/10.3758/BF03206924

Hollingworth, A., & Beck, V. M. (2016). Memory-based attention capture when multiple items are maintained in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance*, 42(7), 911–917. https://doi.org/10.1037/xhp0000230

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics*, 72(5), 1267–1282. https://doi.org/10.3758/APP.72.5.1267

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology. General*, 118(3), 258–279. https://doi.org/10.1037//0096-3445.118.3.258

Kong, G., Alais, D., & Van der Burg, E. (2016). Competing distractors facilitate visual search in heterogeneous displays. *PLoS One*, 11(8), e0160914. https://doi.org/10.1371/journal.pone.0160914

Kong, G., & Fougnie, D. (2019). Visual search within working memory. *Journal of Experimental Psychology: General*, 148(10), 1688–1700. https://doi.org/10.1037/xge0000555

Kristjánsson, Á, & Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research*, 48(10), 1217–1232. https://doi.org/10.1016/j.visres.2008.02.007

Kuo, B.-C., Rao, A., Lepsien, J., & Nobre, A. C. (2009). Searching for targets within the spatial layout of visual short-term memory. *Journal of Neuroscience*, 29(25), 8032–8038. https://doi.org/10.1523/JNEUROSCI.0952-09.2009

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281. https://doi.org/10.1038/36846

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, 22(6), 657–672. https://doi.org/10.3758/BF03209251

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics*, 58(7), 977–991. https://doi.org/10.3758/BF03206826

Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., & Husain, M. (2019). Neural mechanisms of attending to items in working memory. *Neuroscience & Biobehavioral Reviews*, 101, 1–12. https://doi.org/10.1016/j.neubiorev.2019.03.017

Mayer, J. S., Bittner, R. A., Nikolić, D., Bledowski, C., Goebel, R., & Linden, D. E. J. (2007). Common neural substrates for visual working memory and attention. *NeuroImage*, 36(2), 441–453. https://doi.org/10.1016/j.neuroimage.2007.03.007

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, 319(5869), 1543–1546. https://doi.org/10.1126/science.1150769

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, 6(2), 182–189. https://doi.org/10.1038/nn996

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing information during working memory: Beyond sustained internal attention. *Trends in Cognitive Sciences*, 21(6), 449–461. https://doi.org/10.1016/j.tics.2017.03.010

Nobre, A. C., & Stokes, M. G. (2019). Premembering experience: A hierarchy of time-scales for proactive attention. *Neuron*, 104(1), 132–146. https://doi.org/10.1016/j.neuron.2019.08.030

Noonan, M. P., Adamian, N., Pike, A., Printzlau, F., Crittenden, B. M., & Stokes, M. G. (2016). Distinct mechanisms for distractor suppression and target facilitation. *Journal of Neuroscience*, 36(6), 1797–1807. https://doi.org/10.1523/JNEUROSCI.2133-15.2016

Noonan, M. P., Crittenden, B. M., Jensen, O., & Stokes, M. G. (2018). Selective inhibition of distracting input. *Behavioural Brain Research*, 355, 36–47. https://doi.org/10.1016/j.bbr.2017.10.010

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411–421. https://doi.org/10.1037/0278-7393.28.3.411

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, 142(7), 758–799. https://doi.org/10.1037/bul0000046

Olivers, C. N. L. (2008). Interactions between visual working memory and visual attention. *Frontiers in Bioscience: A Journal and Virtual Library*, 13(13), 1182–1191. https://doi.org/10.2741/2754

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1243–1265. https://doi.org/10.1037/0096-1523.32.5.1243

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, 15(7), 327–334. https://doi.org/10.1016/j.tics.2011.05.004

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, 18(2), 283–328. https://doi.org/10.1162/089976606775093909

Ort, E., Fahrenfort, J. J., ten Cate, T., Eimer, M., & Olivers, C. N. (2019). Humans can efficiently look for but not select multiple visual objects. *ELife*, 8, e49130. https://doi.org/10.7554/eLife.49130

Pashler, H. (1987). Target-distractor discriminability in visual search. *Perception & Psychophysics*, 41(4), 285–292. https://doi.org/10.3758/BF03208228

Pashler, H., & Shiu, L. (1999). Do images involuntarily trigger search? A test of Pillsbury's hypothesis. *Psychonomic Bulletin & Review*, 6(3), 445–448. https://doi.org/10.3758/BF03210833

Rabbitt, P., Cumming, G., & Vyas, S. (1979). Modulation of selective attention by sequential effects in visual search tasks. *Quarterly Journal of Experimental Psychology*, 31(2), 305–317. https://doi.org/10.1080/14640747908400729

Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, 24(1), 19–29. https://doi.org/10.1016/S0896-6273(00)80819-3

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 31(2), 248–261. https://doi.org/10.1037/0096-1523.31.2.248

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue

effect. *Attention, Perception, & Psychophysics*, *78*(7), 1839–1860. https://doi.org/10.3758/s13414-016-1108-5

Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics*, *76*(7), 2136–2157. https://doi.org/10.3758/s13414-014-0633-3

Treisman, A. (1992). Perceiving and re-perceiving objects. *American Psychologist*, *47*(7), 862–875. https://doi.org/10.1037/0003-066X.47.7.862

van Moorselaar, D., Theeuwes, J., & Olivers, C. N. L. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention? *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1450–1464. https://doi.org/10.1037/a0036229

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). PsychologyPress/Erlbaum (UK) Taylor & Francis.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(1), 34–49. https://doi.org/10.1037//0096-1523.18.1.34

Woodman, G. F., & Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition*, *14*(4–8), 808–830. https://doi.org/10.1080/13506280500197397

Woodman, G. F., & Luck, S. J. (2002). Interactions between perception and working memory during visual search. *Journal of Vision*, *2*(7), 732–732. https://doi.org/10.1167/2.7.732

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 363–377. https://doi.org/10.1037/0096-1523.33.2.363

Zhang, B., Liu, S., Doro, M., & Galfano, G. (2018). Attentional guidance from multiple working memory representations: Evidence from Eye Movements. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-32144-4