Vision Research 189 (2021) 104–118

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

A neuro-computational model of visual attention with multiple attentional control sets

Shabnam Novin [a,c], Ali Fallah [a,*], Saeid Rashidi [b], Frederik Beuth [c], Fred H. Hamker [c,*]

[a] Faculty of Biomedical Engineering, Amirkabir University of Technology, Tehran, Iran
[b] Faculty of Medical Sciences & Technologies, Science & Research Branch, Islamic Azad University, Tehran, Iran
[c] Department of Computer Science, Chemnitz University of Technology, 09107 Chemnitz, Germany

ARTICLE INFO

Keywords:
Visual attention
Multiple attentional control settings
Divided attention
Decision making
Neuro-computational model

ABSTRACT

In numerous activities, humans need to attend to multiple sources of visual information at the same time. Although several recent studies support the evidence of this ability, the mechanism of multi-item attentional processing is still a matter of debate and has not been investigated much by previous computational models. Here, we present a neuro-computational model aiming to address specifically the question of how subjects attend to two items that deviate defined by feature and location. We simulate the experiment of Adamo et al. (2010) which required subjects to use two different attentional control sets, each a combination of color and location. The structure of our model is composed of two components "attention" and "decision-making". The important aspect of our model is its dynamic equations that allow us to simulate the time course of processes at a neural level that occur during different stages until a decision is made. We analyze in detail the conditions under which our model matches the behavioral and EEG data from human subjects. Consistent with experimental findings, our model supports the hypothesis of attending to two control settings concurrently. In particular, our model proposes that initially, feature-based attention operates in parallel across the scene, and only in ongoing processing, a selection by the location takes place.

1. Introduction

The brain uses the mechanism of visual attention, selecting the relevant and important items from all of the available information, while ignoring other irrelevant information, to make efficient behavioral responses (Carrasco, 2011). In many everyday activities, people need to monitor multiple sources of information simultaneously (Hüttermann & Memmert, 2017), e.g., while walking, driving, cooking, doing sports. The ability to attend to more than one stimulus simultaneously is called divided attention (Dannhauser et al., 2005).

Many theories, especially the earlier ones, propose that the focus of attention is limited to one area or item (Heinze et al., 1994; LaBerge & Brown, 1989; Posner, Snyder, & Davidson, 1980), which is still - to some degree - investigated (Jans, Peters, & De Weerd, 2010). Some studies, however, provide evidence, under appropriate circumstances, for multiple spatial foci (Awh & Pashler, 2000; Dubois, Hamker, & VanRullen, 2009; Frey et al., 2014; Hamker, 2004; Kawahara & Kumada, 2017; Zirnsak, Beuth, & Hamker, 2011) or for a review, see Jans et al. (2010) cf. Cave, Bush, and Taylor (2010). Further, attention may not be static over longer periods of time, but also sample the space in a 7–12 Hz rhythm (Gaillard et al., 2020; Jia, Liu, Fang, & Luo, 2017).

However, divided attention has often been studied with an attentional control set for a single feature. An attentional control set (ACS) contains representations of parameters that control the goal of attentional selection of a target, such as a particular feature and/or a location (Adamo, Pun, & Ferber, 2010; Folk, Remington, & Johnston, 1992; Grubert & Eimer, 2016).

Adamo, Pun, Pratt, and Ferber (2008) suggested that their observed reaction time pattern resulting from spatial cues that matched the ACS supports the evidence for two ACSs for color and location. In their experiment, subjects were required to respond to a target color only at a particular location, but at the same time to a different target color at another location. Targets were preceded by non-predictable cues which were congruent or incongruent to the target in location, or in color, or in both. The reaction times were shortest when the cue matched the target both in the color and location, and longest when the cue did not match the target both in the color and location. Partially matching cues resulted in reaction times in between. Based on the behavioral reaction times, Adamo et al. (2008) concluded that humans can maintain two ACS for different locations in space.

* Corresponding authors.

https://doi.org/10.1016/j.visres.2021.08.009
Received 29 January 2021; Received in revised form 30 June 2021; Accepted 4 August 2021
Available online 5 November 2021

However, reaction time data is not always conclusive about the underlying neural processes. In a later study, Adamo et al. (2010) repeated almost the same experiment, now combined with ERP measurements. Although, the behavioral results of their second study were consistent with their previous one, their cue N2pc data provides no evidence for two ACS for the conjunction of color and location. They found that all cues that match the color of the ACS capture attention, regardless of their spatial match with the ACS. They also investigated the rapid disengagement hypothesis, according to which all cues elicit attention, but then attention becomes rapidly disengaged for non-matches. As the target N2pc amplitude was affected by the spatial match of the cue to the following target, regardless of their color match with the ACS, their data does not confirm a rapid disengagement of attention. The P3, an ERP component associated with attention employing to the candidate target at late stages, e.g., encoding into working memory (Adamo et al., 2010; Irons & Remington, 2013), showed a differential activation for targets dependent on the match of the cue with the ACS.

Irons, Folk, and Remington (2012) investigated whether simultaneous attentional control settings for multiple colors capture attention, regardless of their location. Subjects were required to identify the target character which appeared in either of two predefined colors along with a distractor character in an irrelevant color in one of 4 locations. The target was preceded with a non-predictable cue. The authors varied the cue to match the color of the target (relevant match), or the color of the second ACS (relevant nonmatch), or having no match with any ACS (irrelevant). Although the reaction time is fastest for cues matching the target, their results show that all cues matching either ACS lead to a significant cueing effect compared to cues that do not match the ACS, which confirms the hypothesis that subjects can use simultaneous attentional control settings for multiple colors. Thus, the study of Irons et al. (2012) partially supported the results of Adamo et al. (2010), but addressed only the aspect of allocating attention to multiple color-based ACSs and not to conjunctions of color and location.

In a later study, Irons and Remington (2013) investigated if two ACSs can be maintained also when they are defined as conjunction of color and location. They designed an experiment with two simultaneous RSVPs of letters at the right and left of fixation. Subjects had to identify a target letter which could be either in green within the left stream or in red within the right stream, while colors varied for the individual letters within the stream. A relevant distractor letter had a target color but was presented at the wrong side, e.g. green in the right stream. They observed that, when a relevant distractor was presented prior to the target, it led to an attentional blink. Their observation suggests that subjects are not able to use two conjunctive ACSs.

The hypothesis of allocating attention to two ACS in two different locations was also not confirmed by Becker, Ravizza and Peltier (2015) and Liu and Jigo (2017). Becker et al. (2015) discussed that the observation of Adamo et al. (2010) about the effect of congruency of cue and target on the reaction times, may not be clear evidence for the simultaneous allocation of two ACS, but rather it could be that the cue primes the color or location of the target or both. Becker et al. (2015) designed a paradigm with 8 locations for the stimuli and two possible targets. They did not use cues to avoid any possible priming effect. Subjects had to identify the target letter which could be either red in the left hemifield or green in the right hemifield. The target was accompanied by distractors which could be relevant or irrelevant to the ACSs. The relevant distractor appeared in the same hemifield of the target and with the color of the other possible target. Their results showed that the relevant distractors decreased the performance significantly more than the irrelevant distractors. Since the relevant distractors were matching one of the two possible targets in color, but not in location, Becker et al. (2015) concluded that the subjects were allocating their attention globally to the two defined colors rather than allocating to two different possible locations for the targets. Liu and Jigo (2017) discussed that there is a limitation in attending to two different colors simultaneously. They used a cueing paradigm showing a circle of noisy colored dots in a fixed location, asking the subjects for color detection. They compared three conditions of presenting no cue, one cue, and two cues by applying different levels of coherence for the target and measuring the detection threshold. Their results showed that the subjects had a better performance in one cue and two cue conditions compared to no cue, but the performance decreased in the two cue compared to the one cue condition. They concluded that attention to two colors is possible, but leads to a cost compared to a single cue condition.

Berggren et al. (2017) explored whether space and color are equal components of an attentional template, such that subjects can search for a particular color in only a part of the visual field and ignore the other. Their task design should ensure that subjects apply the feature-location combination to obtain a good performance. Similar to Adamo et al. (2010), the N2pc amplitude was comparable for both matching and non-matching cues. Thus, attentional capture, measured by the N2pc, can not be selective to only a particular feature in space, even when it would be beneficial for the task performance. Their results support the view that initially, attention is allocated in spatially global fashion to all features (regardless of location) that match the target, but only later, attention narrows down to the full target sets.

As the more recent studies suggest that subjects, during their first allocation of attention, are not able to fully rely on the color/space combination of two ACS, but rather search globally for color only (Adamo, Pun, & Ferber, 2010; Becker, Ravizza, & Peltier, 2015; Berggren, Jenkins, McCants, & Eimer, 2017; Irons & Remington, 2013), the question arises of how such attentive selection develops over time. Further, it is not yet clear how the congruency of cue and target stimulus affects the behavioral responses and the ERP measurements in the experiment of Adamo et al. (2010). In order to better understand the putative underlying neural processes of how ACS may be implemented and affect attention and how cues capture attention, particularly with respect to their match or non-match to properties of the ACS, we designed a neuro-computational model of attention and decision making. Previous models often focused on either feature-based or spatial mechanisms and the neural mechanisms of feature-based attention combined with spatial attention has been rarely studied (Andersen et al., 2011). However, the combination of these two mechanisms is required to explain the debates about the tasks of attending to two features in combination with two locations. The role of two ACSs is also not much investigated in previous computational models.

By means of our neuro-computational model we aim to provide a better understanding of the following questions. When we agree with the more recent evidence that subjects are only able to initially search globally for the two color features, how does then attention develop over time and what are the underlying neural mechanisms of it? Why does the reaction time data of Adamo et al. (2008) has been taken as evidence for two conjunctive ACSs for color and location? Although Adamo et al. (2010) confirmed the reaction time pattern, their ERP data does not support evidence for two conjunctive ACSs. Were the results affected by priming as hypothesized by Berggren et al., 2017? When subjects may initially search globally for color, why does the data of Adamo et al. (2010) show a similar RT for partial matches of the cue to the target, in either location or color, but no imbalance between these two conditions?

We investigate the conditions which allow us to replicate the data of Adamo et al. (2010) and provide predictions about the underlying neural activities. For this purpose, we extended recent modeling work (Beuth & Hamker, 2015; Zirnsak, Beuth, & Hamker, 2011) to allow for multiple attentional task sets, consistent with the idea of a spatially

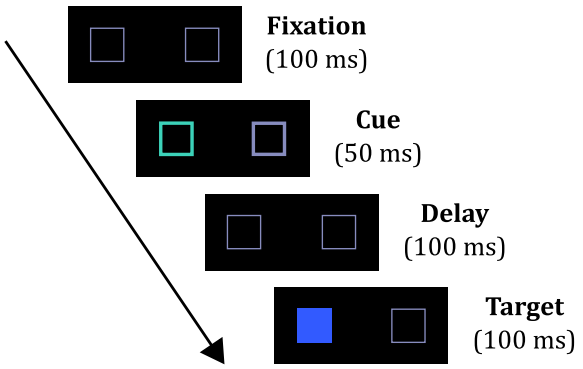

Fig. 1. Stimulus sequence in one example trial, used as input in our model. The images were presented as in the experiment of Adamo et al. (2010). The sequence consisted of the fixation, cue, delay, target image and again the delay until the model's response. The cue could either match or not match the ACS or be a neutral cue. The target was either a blue square presented on the left side or a green square presented on the right side. The stimuli above reflect a S+C- condition as the cue matches the spatial location (S+), but not the color (C-) of the ACS in relation to the target. See main text for the abbreviations and details. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

global feature-based attention template proposed by Berggren et al., 2017, and integrated our model of attention with a simple model of decision making based on the framework of Cohen, Dunbar, and McClelland (1990).

2. Materials and methods

2.1. Experiment description

We simulated the experiment of Adamo et al. (2010), in which the authors ask human subjects to maintain and apply two attentional control sets (ACS) at a time. The subjects are required to respond to two target items jointly defined by a particular color and location, e.g. blue on the left side and green on the right side. Their design is of particular interest for at least two reasons. First, the transfer of attention from the cue to the target is not trivial and the mechanisms are not well explored by neuro-computational models of attention. Second, the relationship between the reaction time and physiological data is not clear. The RT data suggests the possibility of forming two different ACSs to different locations, while the ERP data rejects this possibility consistent with more recent studies (Becker, Ravizza, & Peltier, 2015; Berggren, Jenkins, McCants, & Eimer, 2017).

The experiment was simulated by presenting a stream of images which constitutes the input to our model (Fig. 1). First, a fixation image appeared which consisted of two squares with thin gray borders on black background presented on both sides of the center of the image. In the experiment, the fixation image was presented for 1000 ms and with a fixation cross in the center of the image. We changed for our model the presentation time to 100 ms and omitted the fixation cross, because at 100 ms, the model already reached its baseline activity level, as revealed by control simulations and the model had been predefined to fixate at the image center. Afterwards, a cue image appeared which could either contain a blue, green, or gray (as neutral) cue in form of a colored border at one of the two squares. We used the same input stream as in the experiment, so the cues were of course non-predictive of the targets. Then, a delay image, which was identical to the fixation image, was presented for 100 ms, and finally, the target image appeared containing the blue filled square on left or the green filled square on right. After target presentation, the basic input, which was the fixation stimulus, was shown and we waited until the response exceeds a defined threshold for a maximum of 750 ms, again as in the experiment. A vice versa combination was also performed in the experiment via a secondary participant group to account for subject specific-biases, which are not present in the model. Therefore, we simulated the model for both input streams with either blue on left target stimulus or green on right target stimulus. We averaged the results of both cases to get the final results. All of the images in the stimulus sequence were considered with roughly the same stimulus size and position details as in the experiment. The blue and green color used in the stimuli were chosen to induce equal activity in the color encoding cells of the model. Also the gray color was chosen to induce balanced activities associated to the blue and green color. The human subjects' task was to press a button as quickly and as accurately as possible when they see one of the two defined targets. Similar to the experiment, the correct response in the simulation was to decide for pressing a button when detecting the target stimulus of blue on the left or green on the right. The experiment included also no target trials, where the target was shown in a target color, but at the opposite location, i.e. trials with a blue filled square on right or a green filled square on left, in which subject should not respond. We also simulated no target trials to constrain our model parameters. Here, the model is set to avoid a response. The model parameters are identical in all conditions.

The cue stimuli could be matching or unmatching with respect to the color (C) and spatial location (S) of the targets. Four conditions were considered as in the experiment. The conditions are shown in Fig. 2, for the example of the target of a blue square on the left. Corresponding conditions were considered for the target of a green square on the right in terms of congruency of cue and target stimulus. In condition S+C+, the cue is matching the target in both, in location (spatial match, S+) and color (color match, C+). In condition S+C-, the cue is matching the target in location, but had the opposite color (color non-match, C-). In condition S-C+, the cue is matching the target in color, but was presented at the opposite location (spatial non-match, S-). Finally, in condition S-C-, the cue is not matching the target neither in location and nor in color. The condition of the neutral cue was also simulated in which, the cue contained gray bordered squares in both sides.

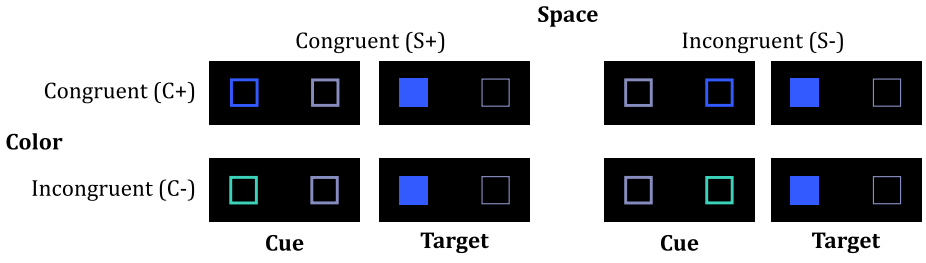

Fig. 2. Conditions considered in the simulation of the study of Adamo et al. (2010). In the experiment, four conditions were considered in relation to the congruency of the cue and the target in terms of color (C) and spatial location (S). In the presented conditions, the target is the blue square on the left. See main text for details about the conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

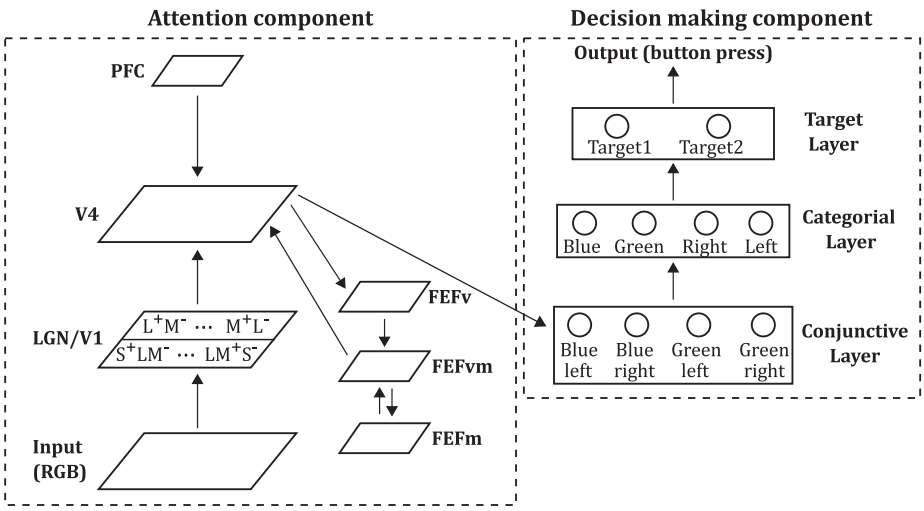

Fig. 3. Diagram of the model. Two components are considered to model the behavior of the subjects, attention selection and decision-making. The attention component is composed of neural layers that represent different areas of the brain engaged in attentional processing, including visual areas LGN/V1 and V4, the prefrontal cortex (PFC) for providing the ACS, and the frontal eye field (FEF) for providing a spatial reentrant signal. The decision-making component receives the information from the cells in the V4 layer of the attention network, which are connected to be activated by the green (blue) square on the right (left) side of the image. It provides the response representing the button press when either target (the blue square on the left or the right square on the right) is shown.

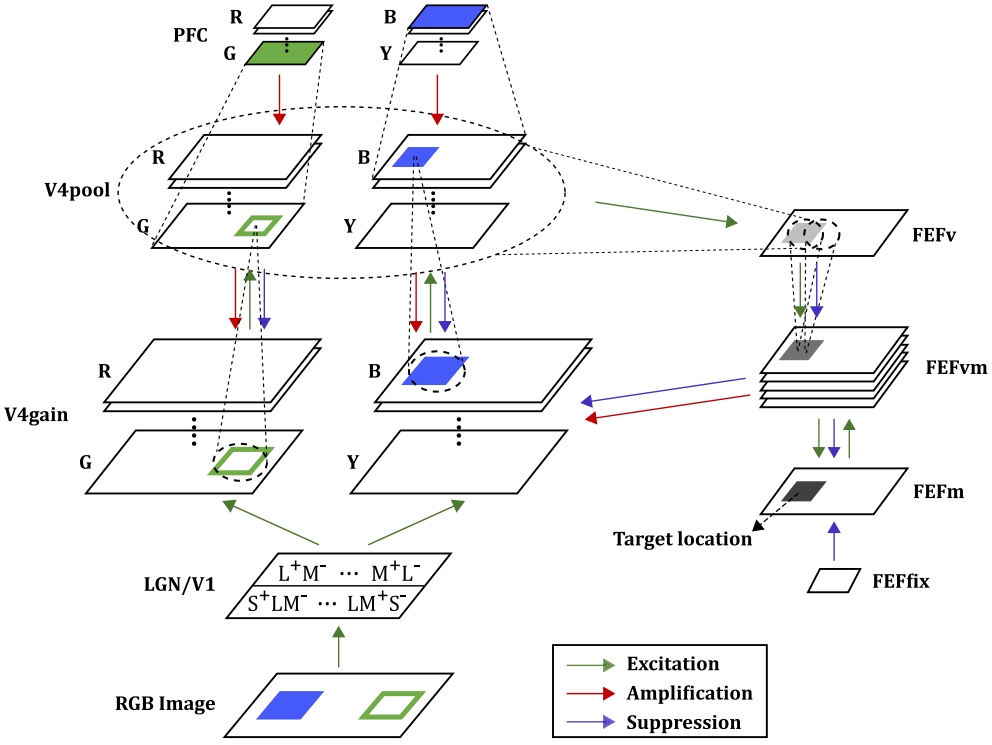

Fig. 4. Diagram of the attention component, illustrated for the condition of S-C- as an example. The cue and target stimulus are shown together in one image just for illustration purpose. The model is composed of the visual areas LGN, V1, and V4gain which, for simplicity, realize only the bottom-up processing of the input image. The visual V4pool receives the excitation signal from V4gain and in parallel, it receives top-down attention signal from PFC which interacts with the bottom-up processing. V4pool sends the activities to FEF, which performs saccade planning. FEF is composed of 3 parts named FEF-visual (FEF-v), FEF-visuomovement (FEF-vm), and FEF-movement (FEF-m) cells, respectively functioning in (i) visual processing, (ii) interface between visual and movement processing, and (iii) movement processing to plan the saccade. The parts simulate the representative cell types in the FEF (Schall, 1991). FEF-vm sends back its activity to V4gain to modulate the response in the visual areas. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. Model overview

The model is composed of two components named “attention” and “decision making” (Fig. 3). The attention component models the dynamics of selection, and the decision-making component models the higher cognitive stage for the behavior of subjects in order to decide about to press the button as a response or not.

2.2.1. The attention component

The attention component (Fig. 4) is a slightly modified and extended version of previous work (Beuth & Hamker, 2015; Zirnsak, Beuth, & Hamker, 2011) and its biological plausibility has been previously demonstrated (Hamker, 2005). We build upon the version of Zirnsak et al. (2011), who proposed in their model reentrant processing from the FEF visuomovement cells being more consistent with data from Cohen et al. (2009) and Ray, Pouget, and Schall (2009), compared to the FEF movement cells earlier proposed by Hamker (2005). The model is composed of rate coded neural layers that represent different areas of the brain engaged in attentional processing. It comprises the visual areas LGN, V1, and V4, the prefrontal cortex (PFC) functioning in executive control, and the frontal eye field (FEF) functioning in saccade planning. We first explain the model regarding the functions of its different areas and then their mathematical description.

The input of the model is a RGB image per time step, providing via LGN/V1 the bottom-up signal which interacts in V4gain and V4pool with top-down attention processing. V4gain receives spatial feedback from the FEF, while V4pool receives the goal-driven or top-down signals from PFC. This layered V4 architecture has been motivated by Beuth and

Hamker, 2015 inspired by the main cortical pathways and is not included in the model version of Zirnsak et al. (2011).

First, the image is processed by lower level visual areas LGN and V1 that encode simple features. We model here only the color features relevant for the task: A red-green (RG) and blue-yellow (BY) color difference channel. The layers are implemented based on the physiological properties of the cells in these brain areas (Gegenfurtner, 2003). The feature-encoding neurons are spatially arranged, constituting retinotopic planes. The color contrasts are computed based on the differences between combined responses of L, M cone cells, and S cone cells, respectively, referring to large, middle, and small wavelength cones.

The core of the model composed of V4gain and V4pool implements the Biased Competition framework (Desimone & Duncan, 1995), where V4gain roughly refers to layer 4 and V4pool to layer 2/3 of area V4. This principal circuit of the model has been demonstrated to be consistent with neural recordings of numerous attention experiments (Beuth & Hamker, 2015) and has been here extended to full images. The neurons in V4gain and V4pool layer are grouped into two channels covering the feature-spaces and arranged in retinotopic planes similar as in V1. The activities of neurons in the V1 layer are sent to neurons of V4gain to be modulated by spatial and feature-selective top-down signals. The neurons of V4pool, which have larger and overlapping receptive fields compared to V4gain neurons, spatially pool the modulated activities of V4gain to increase spatial invariance.

V4pool layer receives a top-down signal from PFC maintaining the target color and sends, in turn, a feature-based feedback to V4gain layer, amplifying the activity of a number of V4gain cells, within its receptive field. In parallel, V4pool neurons project their activities to the FEF layers, which are involved in the planning of eye movements to the target location. As no saccades are required in this task, the FEF here only provides a spatially selective feedback signal to V4. The FEF of the model is composed of 3 parts named FEF-visual (FEF-v), FEF-visuomovement (FEF-vm), and FEF-movement (FEF-m) which are defined based on cell properties of the FEF area in the brain (Schall, 1991; Zirnsak et al., 2011). The activity of FEFv neurons is computed based on the maximum activity over all feature channels of V4pool at each location, and its output represents a kind of saliency map. Due to the top-down target color bias to V4, the saliency map is task-specific. FEFvm cells have some properties of both visual and movement cell types. FEFvm combines FEFv activities in the form of a center (+)-surround (-) type. The FEFvm projects its activities to V4gain (reentrant theory). This recurrent loop modulates the activities in V4gain by exciting the responsive cells to the target location and suppressing the nonresponsive ones, leading to the emergence of spatial attention. In case of eye movements, a second recurrent loop over FEFm iterates until the activities of FEFm neurons exceed a specific threshold and plan a saccade (eye movement). However, here the model is set into fixation mode as required by the task by continuously suppressing the FEFm activity.

2.2.2. The equations of the attention component

We here provide the mathematical description of the model and the equations of each layer. Please also refer to previous models of Beuth and Hamker (2015), Zirnsak, Beuth and Hamker (2011), and Beuth (2019). In each layer, we consider a population of neurons and dynamically update their firing rates. Attention emerges by the dynamics of the recurrent system, so we describe the firing rate changes by differential equations over time. This makes it possible to model the temporal dynamics of attention.

In the following equations, the firing rates of neurons are denoted by $r^{area}_{d,i,x}$. The superscript refers to the associated area and the subscript refers to the neurons' indices. The index x refers to the neuron's location in a topographically organized layer, the index d refers to the associated feature channel (red-green, blue-yellow), the index i refers to the preferred feature in a channel which is associated with the ith neuron in the population. The weights are denoted as follows: $w^{area1-area2}_{x,x'}$ is the connection weight between the neurons of two areas, $w^{area/x}$ refers to spatial-lateral connection weights within an area and $w^{area/i}$ is the connection weight of the cells at a specific location in a specific channel.

2.2.2.1. LGN layer. The LGN layer refers to the processes in the lateral geniculate nucleus (LGN) of the brain and consists of three kinds of cells: Parvo (L-M), Konio (S-(L+M)), and Magno (L+M) that are respectively associated with the processing of Red-Green contrast, Blue-Yellow contrast, and grayscale feature. In order to simulate these cells, the image is transformed from RGB to LMS cone space, and the responses of the L, M, and S cone cells are computed. The L-M cells have four types: two types of ON cells and two types of OFF cells. The ON cells are excited in the center by L/M cones and inhibited in the surround by M/L cones named respectively L+M− and M+L− and the OFF cells vice versa: L−M+, and M−L+ .

2.2.2.2. V1 Layer. Each neuron in V1 has a preference for a specific feature, so different neurons encode different features. The neurons are grouped into two channels. Channel 1 neurons encode the red-green (RG) color contrast and the channel space is arranged between the cells exited by red (L+M−, M−L+) and the cells exited by green (M+L−, L−M+). Similarly, Channel 2 neurons encode the blue-yellow (BY) color contrast and the channel space is arranged between the cells exited by blue (S+LM-) and the cells exited by yellow (LM+S-).

The firing rates of V1 neurons are calculated based on the feature values and the simulated responses of LGN cells by the following equations. The numbers 1, 2 in the indexes of firing rates refer respectively to RG and BY channels and the v values represent the simulated responses of LGN cells.

$$r^{V1}_{d=1,i} = \frac{i-1}{L-1}\left(0.5v^{L+M-} + 0.5v^{M-L+}\right) + \frac{L-i}{L-1}\left(0.5v^{L-M+} + 0.5v^{M+L-}\right) \quad (1)$$

$$r^{V1}_{d=2,i} = \frac{i-1}{L-1}v^{S+LM-} + \frac{L-i}{L-1}v^{LM+S-} \quad (2)$$

where L = 8 is the number of cells in each channel.

2.2.2.3. V4gain layer. V4gain layer receives the activities of the V1 layer and amplifies or suppresses this bottom-up input by spatial information from FEFvm, and feedback from the V4pool layer which in turn receives a feature-based top-down signal from the PFC. The firing rates of V4gain neurons are updated by the following equations.

$$\forall d,i,x : \tau^{V4g}\frac{\partial r^{V4g}_{d,i,x}}{\partial t} = E'_{d,i,x}\cdot A_{d,i,x} - r^{V4g}_{d,i,x}\left(\left(\left(E_{d,i,x}\cdot A_{d,i,x}\right)*w^{V4g/x}\right)*w^{V4g/i}+\sigma\right) \quad (3)$$

$$E_{d,i,x} = 0.6\cdot \max_{x'\in RF(x,V1)}\left(w^{V1-V4g}_{x,x'}\cdot r^{V1}_{d,i,x'}\right) \quad (4)$$

$$\forall d,i,x : \tau^{E'}\frac{\partial E'_{d,i,x}}{\partial t} = E_{d,i,x} - E'_{d,i,x}, \quad \tau^{E'} = \begin{cases} 1, & \partial E'_{d,i,x} \geq 0 \\ 40, & \partial E'_{d,i,x} < 0 \end{cases} \quad (5)$$

$$A_{d,i,x} = 1 + A^{SP}_{d,i,x} + A^{FEAT-V4g}_{d,i,x} \quad (6)$$

$$A^{SP}_{d,i,x} = \max_{x'\in RF(x,FEFvm)}\left(w^{FEFvm-V4g}_{x,x'}\cdot r^{-FEFvm}_{x}\right) \quad (7)$$

$$A^{FEAT-V4g}_{d,i,x} = \max_{x'\in RF(x,V4p)}\left(w^{V4p-V4g}_{x,x'}\cdot r^{V4p}_{d,i,x'}\right) \quad (8)$$

In the equations, the variables E and A denote the effects that a neuron receives, respectively, the excitatory and the amplifying one. The variable E′ represents phasic excitation and the parameters $\tau = 15$ ms and $\sigma = 1.5$ respectively refer to the time constant and attention contrast gain factor. The symbol * denotes convolution.

The excitation to each V4gain cell is obtained by pooling V1 cells firing rates (Eq. 4) and then converting the obtained tonic excitation E to the phasic activation E′ (Eq. 5). Phasic activity is obtained via a delayed decrease of the tonic activity, implemented by a slow time constant for decrease of neuronal activities and a fast time constant for increase (Eq. 5) to mimic parts of the iconic memory and thus to avoid that V4gain activities rapidly decrease to zero during the intervals like the delay stimulus. The amplification signal A is computed as the sum of spatial and feature information of attention (Eq. 6) based on the studies (Saenz, Buracas, & Boynton, 2002). V4gain cells receive spatial feedback from FEFvm (Eq. 7, modeled by 2D-Gaussian weights with $\sigma = 0.83, \sum w = 1$) and feature-based feedback from V4pool cells (Eq. 8, modeled by 2D-Gaussian weights with $\sigma = 2.6$). Furthermore, V4gain cells receive feature-based inhibition from the other cells that encode different features ($w^{V4g/i}$, modeled by 2D-Gaussian weights with $\sigma = 0.4, \sum w = 1$, base inhibition $b = 0.1$), and also they receive spatial inhibition from surrounding cells ($w^{V4g/x}$, modeled by 2D-Gaussian weights with $\sigma = 2.8, \sum w = 1$).

2.2.2.4. V4pool layer. V4pool layer spatially pools the activities of V4gain neurons over each feature, but also receives top-down feature-based attention feedback from the PFC layer. The firing rates of V4pool neurons are updated by the following equations.

$$\forall d,i,x : \tau^{V4p}\frac{\partial r^{V4p}_{d,i,x}}{\partial t} = E_{d,i,x}\cdot A_{d,i,x} - r^{V4p}_{d,i,x}(((E_{d,i,x}\cdot A_{d,i,x}) * w^{V4p/x}) * w^{V4p/i} + \sigma) \quad (9)$$

$$E_{d,i,x} = \max_{x' \in RF(x,V4g)} (w^{V4g-V4p}_{x,x'} \cdot r^{V4g}_{d,i,x'}) \quad (10)$$

$$A_{d,i,x} = 1 + A^{FEAT-V4p}_{d,i,x}, A^{FEAT-V4p}_{d,i,x} = w^{PFC-V4p}\cdot r^{PFC}_{d,i} \quad (11)$$

where E and A respectively refer to excitation and amplification. The parameters $\tau = 15$ *ms* and $\sigma = 0.75$ respectively refer to the time constant and attention contrast gain factor. Excitation is executed by pooling V4gain activities modulated by Gaussian weights (Eq. 10). The amplification signal from the PFC is modeled via one-to-one connections between feature-encoding cells (Eq. 11). V4pool cells also receive the spatial-surround and feature-based self-inhibition respectively via 2D-Gaussian weights $w^{V4p/x}$ and $w^{V4p/i}$ in Eq. 9.

2.2.2.5. PFC layer. The PFC layer sends a feature-based attention signal to V4pool layer, amplifying the activity of those V4pool cells that have responded to the target features. The firing rates of PFC cells are assumed to arrive from a working memory that encodes the target information defined as the expected features of V4pool cells.

2.2.2.6. FEFv layer. FEFv receives its input from V4pool layer in a retinotopic projection, and computes the maximum activity over all feature channels of V4pool at each location and thus, encodes the conspicuity of each location. The firing rates of FEFv neurons are simulated by the following equations.

$$\forall x : \tau^{FEFv}\frac{\partial r^{FEFv}_{x}}{\partial t} = w^{V4g-FEFv}\bullet\frac{1}{\#D}\sum_{d'}\max_{i'} r^{V4g}_{d',i',x} - r^{FEFv}_{x} - w^{FEFv}_{inh}\frac{1}{\#FEFv}\sum_{x'} r^{FEFv}_{x'} \quad (12)$$

where the parameter $\tau = 10$ ms refers to the time constant and $\#D$ refers to the number of channels in V4gain.

2.2.2.7. FEFvm layer. FEFvm cells have some properties of both visual and movement cell types. They get activated by visual information, but also encode target location information to initiate the saccade. In the model, the cells are organized to encode both visual and movement effects. FEFvm receives feedforward excitatory and surround inhibitory drive from FEFv ($r^{FEFv}_{x'}$ in Eq. 13) and the saccade information from FEFm (r^{FEFm}_{x} in Eq. 13). The FEFvm activities project back to V4gain, providing a dynamic spatial attention signal. The firing rates of FEFvm neurons are updated by the following equations.

$$\forall j,x : \tau^{FEFvm}\frac{\partial r^{FEFvm}_{j,x}}{\partial t} = \sum_{x'} w^{FEFv-FEFvm}_{x,x'j} r^{FEFv}_{x'} + w^{FEFm-FEFvm}_{j} r^{FEFm}_{x} - r^{FEFvm}_{j,x} - w^{FEFvm}_{inh}\sum_{x'} r^{-FEFvm}_{x'} \quad (13)$$

$$\forall x, x', j : w^{FEFv-FEFvm}_{x,x'j} = g(x, x', a^{+}\cdot s^{j-1}, \sigma^{+}_{1}) \quad (14)$$

$$g(x, x', a, \sigma) = a\cdot exp\left(-\frac{(x_1 - x'_1)^2}{2\sigma_1} + \frac{(x_2 - x'_2)^2}{2\sigma_2}\right) \quad (15)$$

where the parameter $\tau = 10\,ms$ refers to the time constant and the function g represents a two-dimensional Gaussian function. Depending on j, the Difference of Gaussian function has different positive peak values, implemented by the factor s^{j-1}.

2.2.2.8. FEFm layer. FEFm cells encode target location information by a competition among locations, which is implemented by the following equations. The activity of FEFm cells is formed by a local excitation signal from FEFvm and a global long-range inhibition.

$$\forall x : \tau^{FEFm}\frac{\partial r^{FEFm}_{x}}{\partial t} = w^{FEFvm-FEFm}\sum_{x'} r^{-FEFvm}_{x'} + \sum_{x'} w^{FEFm/x} r^{FEFm}_{x'} - r^{FEFm}_{x} - \frac{1}{\#FEFm}\sum_{x'} r^{-FEFvm}_{x'} - w^{FEFm}_{inh}\sum_{x'} r^{FEFm}_{x'} - w^{Fix} r^{Fix} \quad (16)$$

where the parameter $\tau = 10\,ms$ refers to the time constant. r^{Fix} refers to the fixation cells of the FEF (Hamker, 2005) that prevent the execution of saccades by suppressing the FEFm activity during the time when saccades are not allowed. The level of suppressing is adjusted by the parameter $w^{Fix} = 3$.

The competition among the locations is continued until FEFm cells firing rates exceed a defined threshold at time point t_0, which is considered as the occurrence of a saccade. Due to the inhibition of the movement cells in the current experiment, saccades are not triggered.

2.2.3. The decision-making component

The decision-making component of the model is based on the framework proposed by Cohen et al. (1990) to model the Stroop task. This part is a simple extension to the model of attention and serves the task to map stimuli onto a decision. In these networks, the ACSs with respect to the decision are hard coded in the connectivity. Using the proposed network, we implement the process of accumulating the information over time. The network (Fig. 5) consists of four layers, a

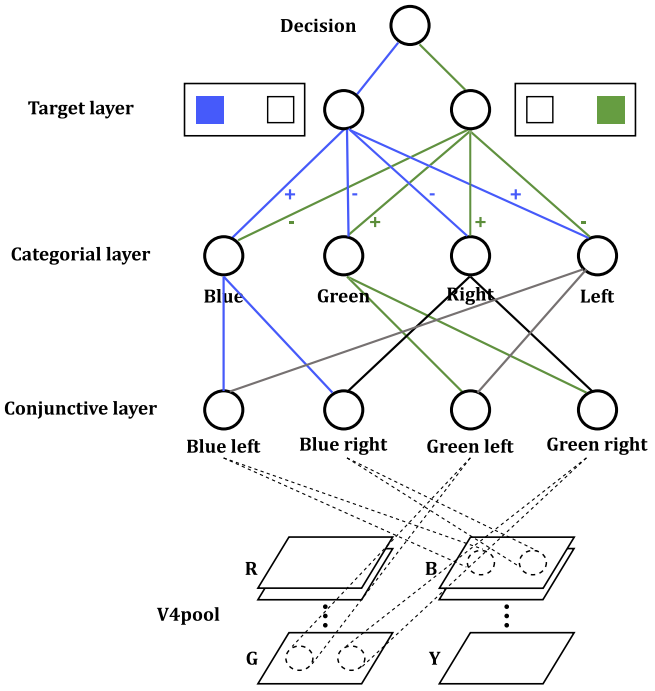

Fig. 5. Diagram of the decision-making component of the model. The decision network receives the information from the cells in the RG and BY channels of V4pool layer of the attention network, to compute activations of conjunction units such as green (blue) square on the right (left) side of the image. The categorial layer separates color from position leading to simple categorial responses. The target layer is composed of 2 neurons representing the two possible targets, one for the blue square on the left and the other for the green square on the right, illustrated by the two rectangles next to the neural units in the figure. The decision layer computes the final behavioral decision, e.g. button press. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

conjunctive layer to converge the information from the attention component, a categorial layer to combine the information, a target layer to identify the targets, and a decision layer which computes the final behavioral decision. The processing in the network is performed by spreading activation in the network through the connections between the units.

The activities from the cells in the V4pool layer of the attention network are spatially pooled to obtain neurons selective to green (blue) on the right (left) side of the image. The categorial layer computes categorial activation in 4 units representing the blue and green color information combined with the right and left location information. Each of the units in the categorial layer projects to both of the units in the target layer. The target layer consists of 2 units, where each one accumulates the evidence for seeing one of the two possible targets. Finally, the evidence for blue on left and green on right are compared with a response threshold in the decision layer. When the activation of one of the target units exceeds the threshold, the decision about pressing the button is made and the associated time is reported as the "decision" time, but also, we use the word reaction time, when we compare the model with the data, taking the assumption that there is no systematic variability in the motor response preparation across different conditions. The response threshold is defined so that its value allows to make correct responses in the target present condition (button-press) and the no target condition (no button-press). In order to mimic the variability among the subjects for making the decision to press the button, and also the inter-subject variability, we applied a variable decision threshold chosen randomly from a Gaussian distribution ($\sigma = 0.1$).

The units in the categorial layer are connected to the units in the target layer in a way that each of the response units receives excitatory input from the units having matching information and inhibitory input from the units having unmatching information (Fig. 5).

Similar to the equations in the attention component, we update the activation of the neurons of the higher-level cognitive stage of decision-making. In categorical and target layer, the activation of the units is calculated based on the weighted sum of the activations from the previous layer (Eq. 17) and then applying a moving average over time, using the Eqs. 18 and 19 respectively for categorial layer and target layer.

$$act_j(t) = \sum_i act_i(t) w_{ij} \tag{17}$$

where act_j refers to the activation of a specific unit j in the layer n that receives the signals from the unit i with activation act_i in layer $n-1$ and w_{ij} is the connection weight between the two units.

$$\text{categorial layer} : \bar{act}_j\,(t) = \bar{act}_j\,(t-1) + \tau^*(act_j(t) - \bar{act}_j\,(t-1)) \tag{18}$$

$$\text{target layer} : \bar{act}_j\,(t) = \bar{act}_j\,(t-1) + \tau^* act_j(t) \tag{19}$$

where $\bar{act}_j\,(t)$ is the time average of the activation in unit j at time t, $act_j(t)$ is the activation in unit j at time t, and τ is the time constant. The time course of processing in the network is provided by Eqs. 18 and 19 which is similar to the differential equations used in the attention component.

A logistic function (Eq. 20) is applied for the categorical layer on Eq. 18 similar to Cohen et al. (1990) to limit the response by a nonlinearity.

$$output_j(t) = \frac{1}{1 + e^{-k^* \bar{act}_j(t)}} \tag{20}$$

where $output_j(t)$ is the calculated output of unit j at time step t and $\bar{act}_j\,(t)$ is obtained by Eq. 18. k refers to gain factor.

2.2.4. Choice of parameters

The decision threshold and the parameters in the decision making network (the weights between the categorial layer and target layer) and the parameters in the attention network (the time constant for the phasic activity in V4gain layer and the strength of the feedback signal sent from FEFvm to V4gain) were set to respond correctly for target present trials and for no target trials (no button press). To meet these conditions, we tried different sets of weights in the decision making network (Fig. 5) and observed that setting the equal weights for color and location connecting the categorial layer to the target layer (Fig. 5) lead to best results. This indicates that subjects may consider color and location equally for their final decision. Too large values for color makes the model being too sensitive in the no target condition as here the correct color is shown at an incorrect location. Similarly, large values for location lead to difficulties with condition 4 in no target case, as the cue fully matches the second defined target. This further justifies the balanced parameters for color and location.

Further, the sum of inhibitory weights is set to be equal to the sum of excitatory weights in the connections between categorical layer and target layer (Fig. 5) to avoid fast but steady accumulation of target evidence in no target trials. Contrary, too weak activation does not allow to accumulate evidence in correct target trials.

The parameters gain factor and time constant in the equations of decision network (Eqs. 18, 19 and 20) were set so, that a variable decision threshold does not lead to very extreme response times. In addition to setting the parameters of decision making network, we also set the strength of the feedback signal sent from FEFvm to V4gain, so that we do not have too small or too big effects for the location information in the decision making network. Further, we set the time constant for the phasic activity in V4gain layer, so that the values of the responses in no target trials have enough difference to the correct target trials that

allows us to separate them by a response threshold.

3. Results

We used almost the same stimuli and conditions as those in the paradigm of Adamo et al. (2010), as illustrated in Figs. 1 and 2. Our model has been designed so that the activities of the layers in the attention network change over time and finally settle to the activities representing the attended feature and location. The final behavioral output of the model is the decision time representing the time of pressing the response button. The calculated decision times are averaged between both target stimuli of blue on left and green on right.

Adamo et al. (2008) concluded from their behavioral data that humans can maintain two spatially distinct ACSs. Their logic was that full attentional capture occurs for S+C+ trials and thus speeds up the response time, while S-C+ and S+C- lead only to a partial match and thus should show similar response times, as observed in their data. Further, they did not expect to see a difference between S+C+ and S+C- trials when subjects cannot form conjunctive ACSs.

As shown in Fig. 6a, our model however predicts rather similar decision times for S+C+ and S+C- trials, although we tried to tune the model parameters to see a stronger difference between these conditions. Thus, our model clearly deviates from the reaction time data of Adamo et al. (2008) and Adamo et al. (2010). The experimental mean reaction times (Fig. 1 in Adamo et al. (2010)) depend on the conditions of how the cue matches the ACS in the order S+C+ < S+C- < S-C+ < S-C- with the values of 445 ms (±25 ms), 490 ms (±20 ms), 500 ms (±18 ms), and 519 ms (±18 ms) respectively, where the RT for S+C- and S-C + are not significantly different from each other (the values in the parentheses indicate the standard error in the related condition). When we directly compare the model response data to the experimental data (Fig. 6c,e) we see a particularly strong deviation in the S+C+ condition. The deviation of the model RT to the subject RT, however, makes sense, because our model searches globally for color only, regardless to their location. Thus, it behaves as expected by Adamo et al. (2008). Importantly, Adamo et al. (2010) found the same type of reaction time distribution like Adamo et al. (2008), but their ERP data clearly showed a Cue N2pc for all cues

Table 1
Comparing the RT results of different integration approaches. The results are averaged between both target stimuli of blue on left and green on right and the values in the parentheses indicate the standard error in the related condition. The bolded row shows the results that made the best fit to the experimental RTs.

	RT in four conditions of the experiment			
Integration approach	S+C+	S+C-	S-C+	S-C-
From target onset	122.1 ms (±3 ms)	130.5 ms (±2.8 ms)	137.8 ms (±3.1 ms)	153.3 ms (±3.3 ms)
From 25 ms before target onset	109.7 ms (±3.2 ms)	130.6 ms (±2.9 ms)	137.4 ms (±3.2 ms)	153.7 ms (±3.4 ms)
From 50 ms before target onset	**92 ms (±3.3 ms)**	**130.7 ms (±2.8 ms)**	**135.4 ms (±3 ms)**	**153.8 ms (±3.4 ms)**

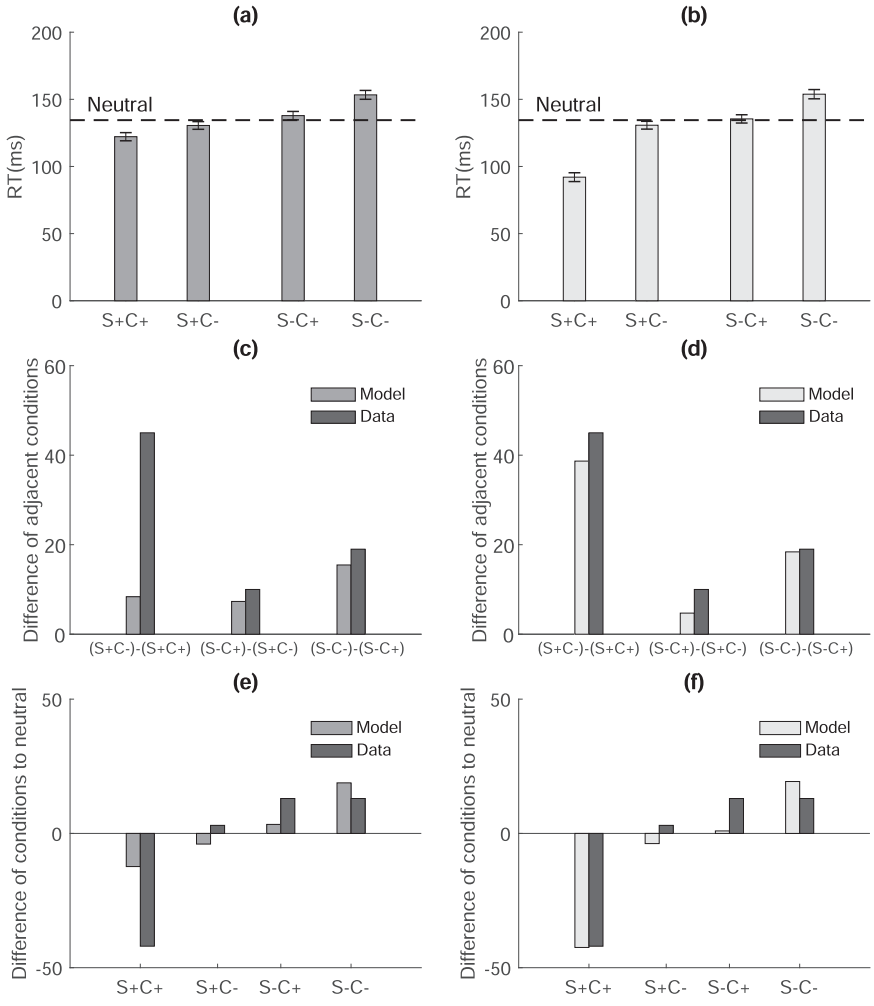

Fig. 6. Model mean decision (reaction) times in the four conditions of the experiment using the approach of accumulating the information (a) from target onset and (b) 50 ms before target onset. The error bar in each condition represents the standard error of the mean in that condition. The dashed line shows the reaction time in the neutral, no cue condition. The model decision times have been computed based on 1000 trials of a variable decision threshold in each condition and stand for the raw decision times as we do not consider motor preparation and execution. Responses are averaged between both target stimuli of blue on left and green on right. The difference between RTs of conditions to each other and to neutral RT in our results compared to data in Adamo et al. (2010) is shown in (c) and (e) respectively using the approach of accumulating the information from target onset and in (d) and (f) respectively using the approach of 50 ms before target onset.

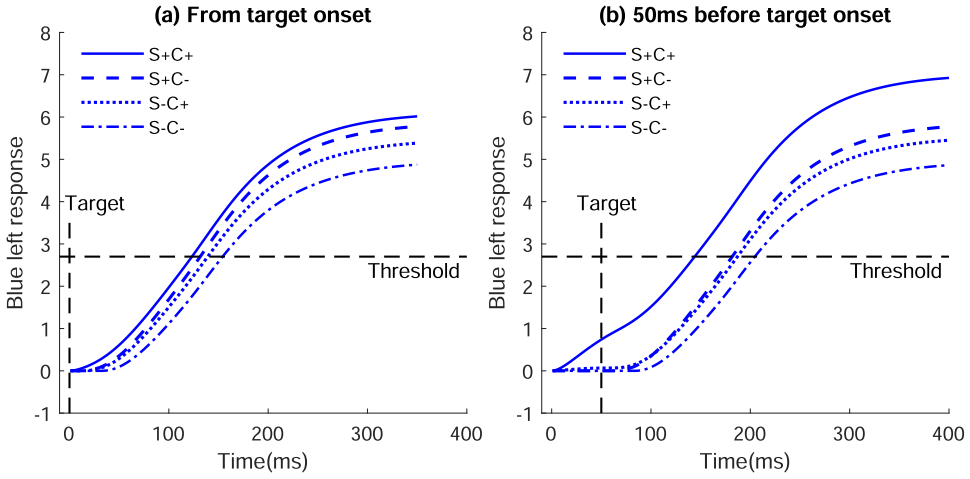

Fig. 7. Comparing the accumulation of the evidence in the target layer between two cases of (a) accumulating the information from target onset vs. (b) accumulating from 50 ms before target onset. The figure illustrates the response of blue on left for the 4 conditions, while the blue on left target stimulus is presented. The arrangement shows when the accumulated evidence in the target layer in each of 4 conditions reaches the decision threshold, which determines the reaction time. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and not just for the one that matches the ACS. Thus, there is obviously a conflict between behavioral and physiological data. As no parameter tuning could make the model match to the behavioral data, we tested if decision priming could explain the dissociation. The task requires human subjects responding to features that are already part of the cue stimulus (color and location). Although subjects were instructed not to respond to the cue, they may already start the decision making process on the features of the cue. In the model results presented in Fig. 6a, we started integration upon target onset. When we start 50 ms prior to target onset we see a clear shift of the reaction time pattern closer to the experimental data (Fig. 6b), while no other parameters have been changed. This improved match with the data is also visible in Fig. 6d,f that shows the difference between RTs of conditions to each other and to the neutral RT in our results compared to data in Adamo et al. (2010). The results, which are averaged between both target stimuli of blue on left and green on right, are presented in Table 1. The results in Table 1 are shown also for the integration from 25 ms before target onset to compare the different starting times for integration of information.

Fig. 7 shows that the particular choice of the threshold does not determine our result. If we accumulate the information from target onset (Fig. 7a), we can not find a good threshold value to strongly separate the S+C- from the S+C+ condition as observed in the data. While when we accumulate the information from 50 ms before target onset (Fig. 7b), the obtained RT pattern fits well to those from human subjects. Models with different parameters do not change this fundamental model prediction.

We now look in more detail how the model, in the condition of integration of 50 ms prior to target onset, predicts the development of attention in the four different conditions. Therefore, we first have a look at the neuronal responses in the visual cortex, more precisely in visual area V4. Fig. 8 visualizes the activation of selective neurons in V4pool, from fixation onset until the end of the simulation in a typical trial of each of the four conditions. The neural activities are shown for the case that the target stimulus is the blue square on the left. We depict those neurons, whose preferred stimuli are part of the ACS either in color or location. The neural activities are boosted in conditions, where the cue or target matches the attentional template (blue and green). The gray colored stimuli, which is a mixture of RGB colors, also induces some activities in the neurons representing green/blue. The response to the cue is not spatially selective, but reflects the match with the target color template only (compare condition S+C+ vs. S-C+ and S+C- vs. S-C-), as the attentional top-down signal acts globally on the whole input, i.e. independent from the spatial location of the stimulus, PFC increases the gain of all blue and green encoding cells in their entire visual field. This is consistent with the fMRI evidence provided by Serences and Boynton (2007) that feature-based attention acts globally on the entire visual field. Eimer (2014) also discusses that the template signals in PFC are mostly location independent. During presentation of the delay image, all of the activities decrease to some extent. The response to the target stimulus (here, the image including the blue square on the left) develops from different starting conditions triggered by the cue and thus, affects the reaction times differently in the four conditions.

The activities of the attention component are passed to the decision-making component. The conjunctive layer pools the activities from V4pool neurons in a way to encode the green (blue) square on the right (left) side of the image, which implements parts of the rule to respond as required.

Fig. 9 shows the activation of the neurons in the categorial layer which integrates the information of the conjunctive layer to compute a categorical activation in the 4 units representing the blue and green color information and also right and left location information. The same explanations of Fig. 8 apply here. Initially, the left and right activities indicate on which side a cue matches the target color. Since there is no difference between the conditions with a spatial matching or non-matching cue to the ACSs, attention is not allocated to two different colors in two different locations, but rather, attention is allocated globally to the two colors of blue and green. However, later after cue onset, the difference between the four conditions becomes different as either the most active color or location switches in the S+C-, S-C+, and S-C-, but not in the S+C+ condition. In the conditions S+C-, S-C+ and S-C- compared to S+C+, the non-target information is more active and acts as distracting neuronal information to identify the target stimulus. This leads to slower accumulation of the information in the response unit associated with the presented target and resulting in slower response time, particularly in condition S+C-, S-C+, and S-C-.

Color and location information become combined in the target layer for accumulation of evidence for each target (Fig. 7). Possible negative evidence is set to zero during the accumulation. When the accumulated evidence for one of the two defined targets exceeds a threshold, the selection time for a button press is registered.

The similarity of the RTs in S+C+ and S+C- conditions in case of no decision priming (Fig. 6a and Fig. 7a) can be explained as follows. In condition S+C-, since the color of the cue matches the color of one of the ACSs, the activity of V4pool cells increases by receiving top-down signal from PFC, and the increased activities are sent to FEFv layer. This causes

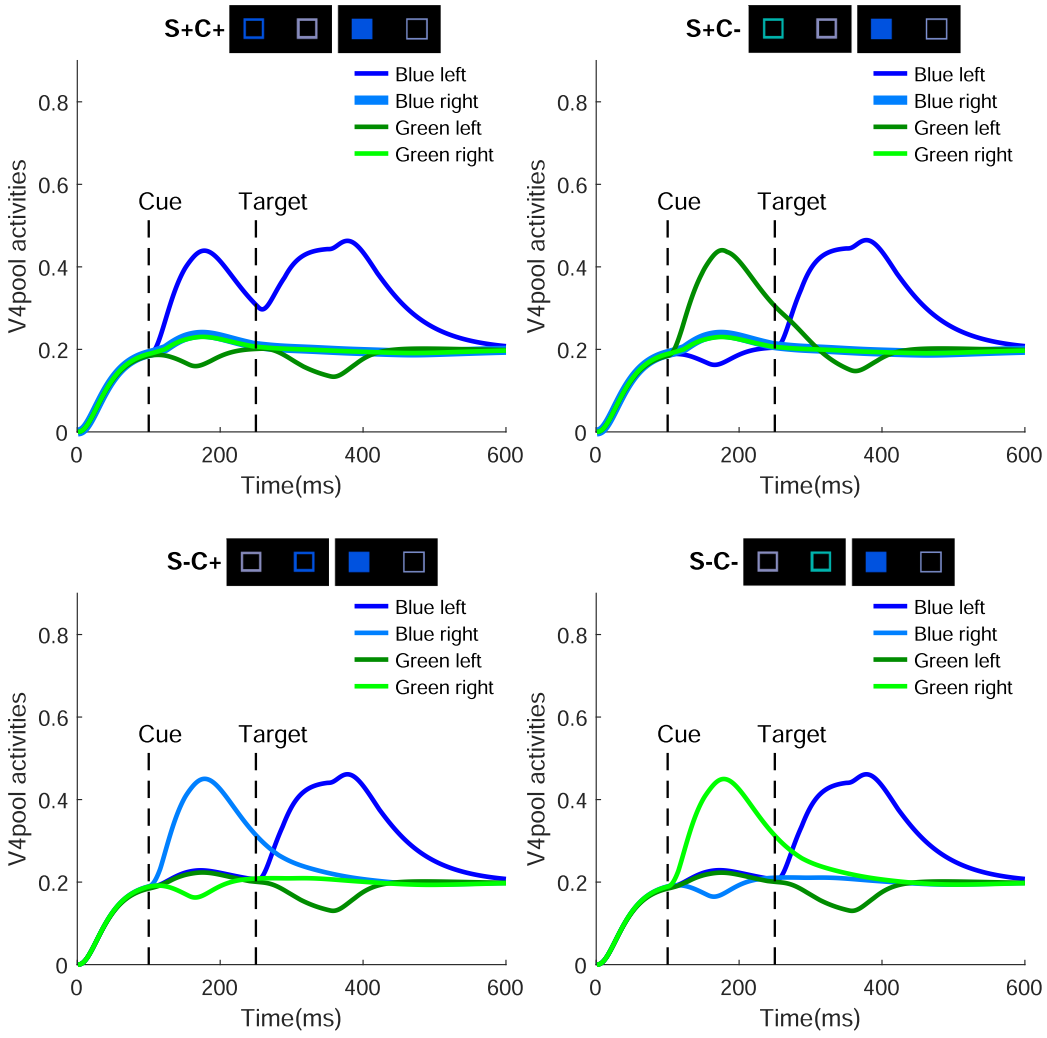

Fig. 8. The course of neural activities in the highest visual layer of the model, V4pool from fixation onset until the end of the simulation in four conditions. The neural activity plotted in different colors is obtained from 4 individual model neurons which match the preferred color and location indicated in the legend. The dashed lines show the onset time of cue and target stimuli. The conditions are presented on top of the plots, including the cue and target image per condition. The activities are shown for the case that the target stimulus was the blue square on the left. The conditions S+C+, S+C-, S-C+, and S-C- refer respectively to the conditions that the presented cue image included: (i) the blue color around left square, (ii) the green color around left square, (iii) the blue color around right square, and (iv) the green color around right square. Note, that the responses for the neurons encoding blue on the right and green on the right are almost identical in the top-panel, hence, we have illustrated the activity for blue on right using increased line width to differentiate these two activities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that FEFvm layer sends stronger reentrant signals to V4gain in the location of the activated color which would be the left location in this case. Please note, that the FEF pools over feature space and thus its responses are invariant to feature identity. In consequence, the presented target benefits from the cue induced attentional capture effect, although the cue met the defined target color but not the defined target location.

S+C+ and S+C- will not lead to identical responses, due to an additional object/feature priming effect, as the cue in S+C+ activates overlapping neuron populations with the target stimulus (Fig. 8).

The similarity of the RTs in S+C– and S–C+ conditions can be explained in a similar way. In condition S+C-, since the color of the cue matches the color of one of the ACSs, the activity of V4pool cells increases by receiving top-down signal from PFC, and the increased activities are sent to FEFv layer. This causes that FEFvm layer sends stronger reentrant signals to V4gain in the location of the activated color which would be the left location in this case. In consequence, more left information will be accumulated in the decision network which increases the location-related excitatory influence to “blue in left” response unit. However, since the cue does not match the color of the correct response (as the target stimulus in this case is blue in left), it increases the color-related inhibitory influence to “blue in left” response unit. In condition S-C+, the color of the cue matches the color of the ACS, but the cue is presented at the wrong location. Since the cue matches the color of the target, it increases the color-related excitatory influence to “blue” response unit, which is a form of feature/object priming. Therefore, in both conditions of S+C- and S-C+, the sum of excitatory and inhibitory influences to the response units is almost equal which leads to similar RTs in these two conditions.

We also investigated how the RT results are affected by FEFvm reentrant signals, which implements attentional capture in the model. Fig. 10 shows the results when we integrate the information from target onset and deactivate FEFvm reentrant signals. Without any decision priming and without attentional capture one would expect rather equal responses. However, the benefit of the S+C+ condition can be traced

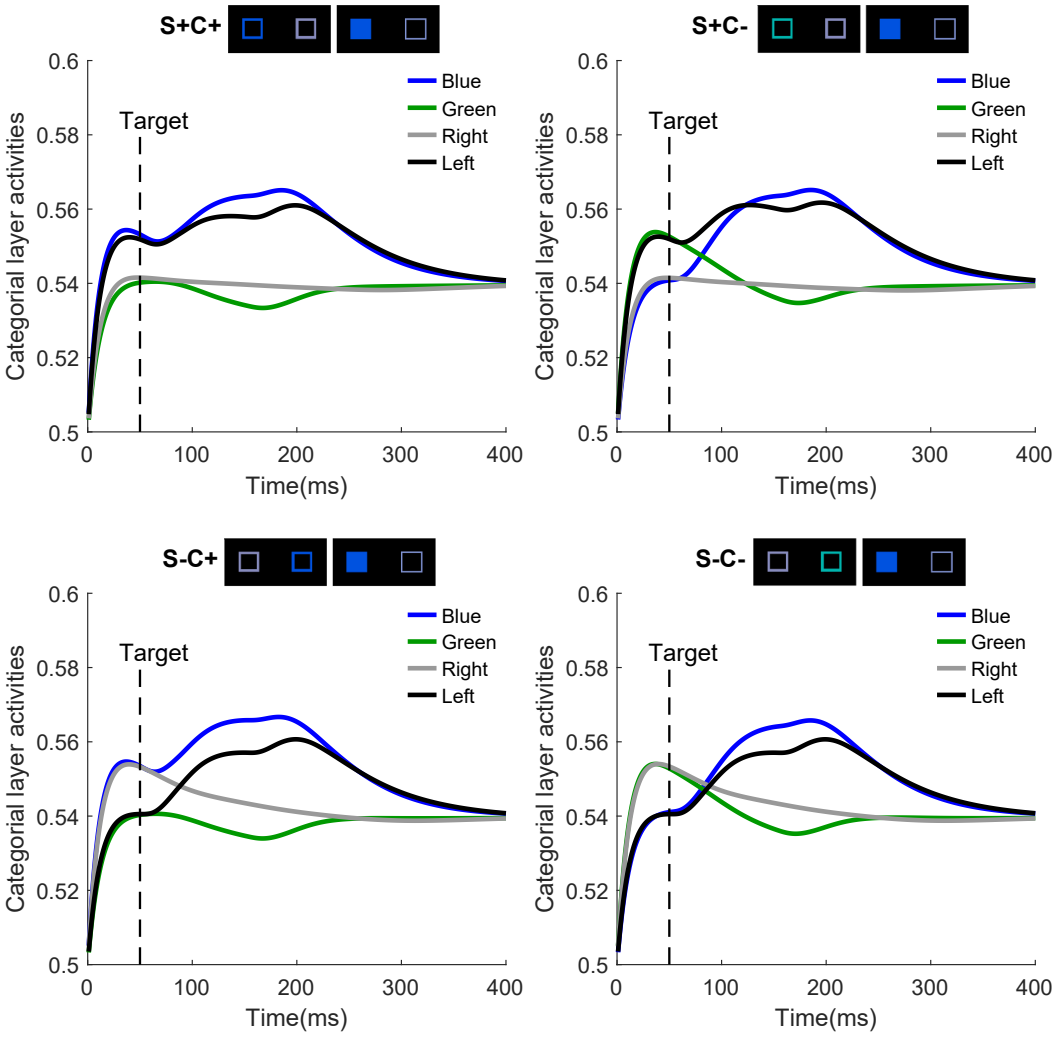

Fig. 9. The course of neural activities in the categorial layer of the decision-making network for the categories blue, green, right, and left, from 50 ms before target onset until the end of the simulation in four conditions. The dashed line shows the onset time of target stimuli. The activities are shown for the case that the target stimulus was the blue square on the left. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

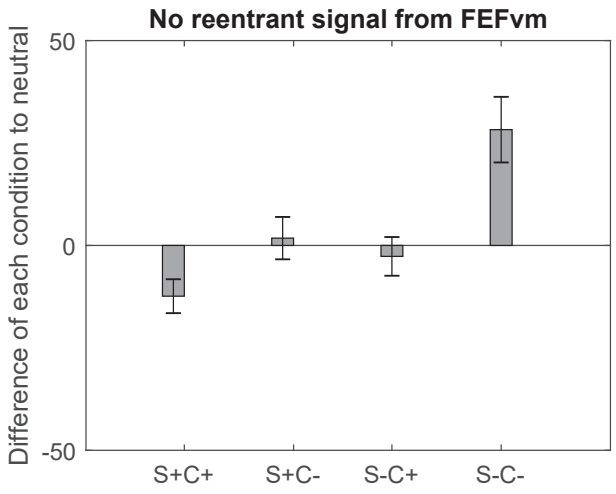

Fig. 10. The difference between RTs of four conditions to neutral RT, where FEFvm reentrant signals are deactivated and the information is accumulated from target onset. The error bar in each condition represents the standard error of the mean in that condition. Responses are averaged between both target stimuli of blue on left and green on right.

back to a feature/object and location priming benefit (Fig. 9) while the S-C- condition suffers as the cue does neither prime location nor color.

In summary, we have explained under which conditions our model replicates the experimentally observed behavioral data. In order to know how the attentional processing is affected by the attentional control sets and the congruency of cues and targets, we now compare our model activity to ERP signals of Adamo et al. (2010). We chose V4pool layer of the model as a comparison to activities of ERP signals recorded from parietal, parieto-occipital, and occipital channels. The V4pool layer of the model simulates a part of the brain area V4 in the visual cortex, located in the occipital lobe. V4 might not be a complete match to the recording sites, but still gives us a good hint to compare the model with ERP dynamics and provide some knowledge about the incidents in ERP signals.

The Cue N2pc recorded by Adamo et al. (2010) aims at testing in how far a cue that matches the attentional control set (ACS) in location triggers spatially selective attention (good cue) compared to a non-matching cue (bad cue). In order to compare the results of the model with Cue N2pc signals, we computed the average of mean V4pool activities of all model neurons (not only the selected ones plotted for visualization purposes previously) for the conditions illustrated in Fig. 11 which resembles the averaging effect of the ERP measurements in the brain. The presented activities are averaged between the results of

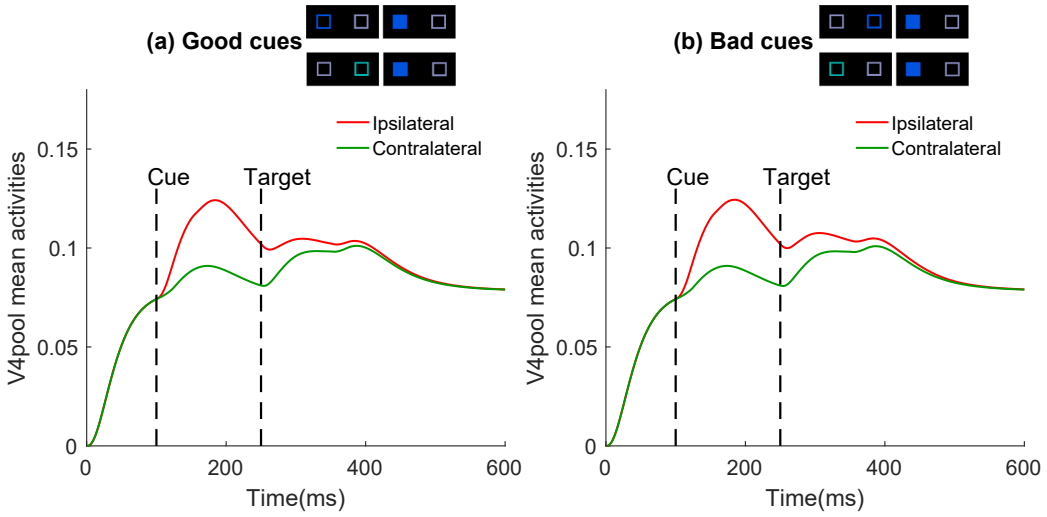

Fig. 11. V4pool mean activity at the ipsilateral side and contralateral side of the cue, shown for good cues (a) and bad cues (b) from fixation onset until the end of the simulation. The activities are averaged between S+C+ and S-C- conditions to represent the activities for good cues and between S+C- and S-C+ conditions to represent the activities for bad cues. The dashed lines show the onset time of cue and target stimuli. The presented activities are averaged between the results of blue on left and green on right target stimuli, where the schematic figure beside the titles of the figures illustrates the conditions for the example case of blue on left target stimulus. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

using blue on left target stimulus and green on right target stimulus. We computed the difference between ipsilateral and contralateral activities to obtain a comparable activity to Cue N2pc (Fig. 12) which represents the difference between the ipsilateral and contralateral brain hemisphere relative to the attended location. When we compare good cues and bad cues, both show a Cue N2pc with similar amplitude and latency, which is consistent with the experimental results where no significant differences have been found between both cues (Fig. 2 in Adamo et al. (2010)). This shows that any cue triggers an attentional shift. The good cues match one of two ACSs both in color and location. The bad cues match one of two ACSs only in color, but not in location. Therefore, the similarity of the results for good and bad cues demonstrates that initially, attention is allocated in spatially global way to the cues based on their color matching with the ACSs, irrespective the match of their location with ACS. In the model, this is reflected by feature-based attention to color.

The Target N2pc recorded by Adamo et al. (2010) reveals the amount of spatial attention triggered after target presentation. They categorized Target N2pc recordings into four conditions of S+C+, S+C-, S-C+, S-C- referring to the congruency of the cue to the ACS, as the following target stimulus always matches the ACS in the trials considered (Fig. 3 in Adamo et al. (2010)). We computed a model Target N2pc, by taking the difference between ipsilateral and contralateral mean V4pool activity in each of the four conditions (Fig. 13). The presented activities are averaged between the results of using blue on left target stimulus and green on right target stimulus.

After presenting the target (which is either blue on left or green on right), the activity in the location of the target develops differently in the four conditions dependent on the congruency of cue and target: Consistent with the spatial congruency effect reported in the

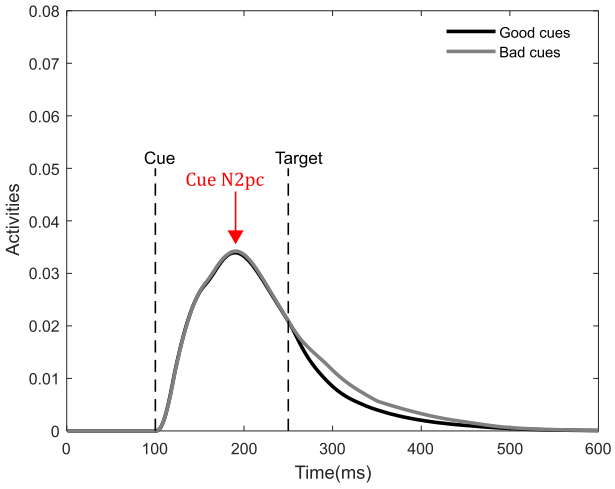

Fig. 12. Difference of V4pool mean activity at the location of right and left stimulus compared between good cues and bad cues from fixation onset until the end of the simulation. This difference activity is a comparable measurement to the N2pc. Good and bad cues show a similar difference activity, albeit the slight distance between both curves. The dashed lines show the onset time of cue and target stimuli.

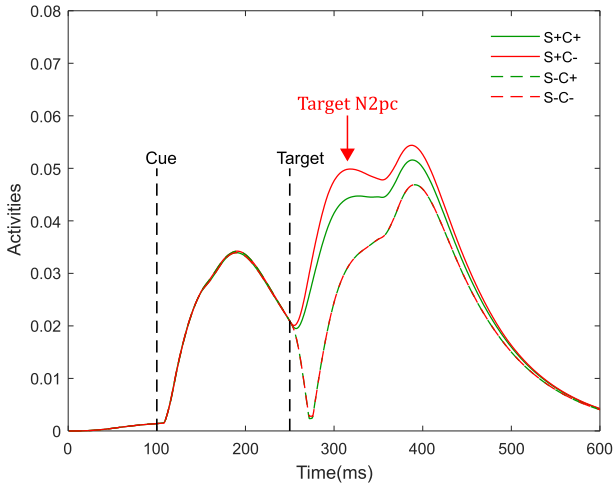

Fig. 13. Difference of V4pool mean activity at the location of right and left stimulus, shown for all four conditions of S+C+, S+C-, S-C+, and S-C- from fixation onset until the end of the simulation. This difference activity is again a comparable measurement to the N2pc. The dashed lines show the onset time of cue and target stimuli. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

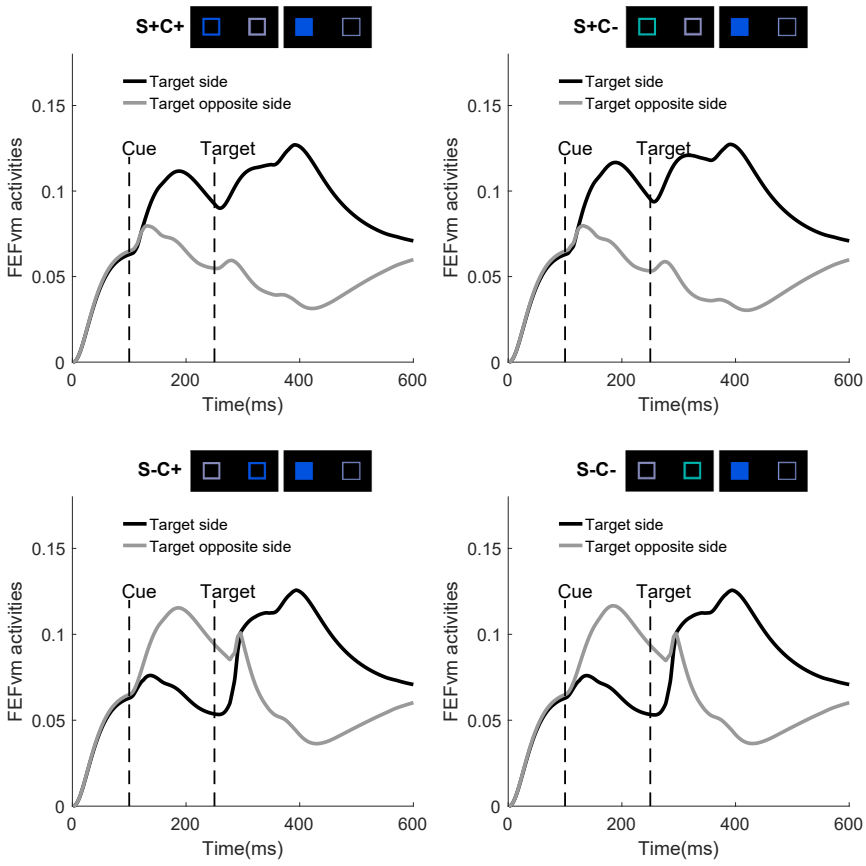

Fig. 14. FEFvm mean activity at target side (left) and target opposite side (right), shown for all four conditions of S+C+, S+C-, S-C+, and S-C- from fixation onset until the end of the simulation. The FEFvm activities are an indicator for a reentrant signal from the FEF. The activities are shown for the case that the target stimulus is blue on left. The dashed lines show the onset time of cue and target stimuli. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

experiment, the peaks appearing after target onset, occur faster in the conditions when the target and the preceding cue are spatially congruent (S+C+ and S+C-) compared to the spatially incongruent conditions (S-C+ and S-C-). Also, consistent with the color congruency effect reported in the experiment, in color congruent conditions, the peaks occur slower (compare S+C+ vs S+C- and S-C+ vs S-C-). In terms of the peak amplitudes, spatially congruent conditions (S+C+ and S+C-) have a higher amplitude compared to spatially incongruent conditions (S-C+ and S-C-) which is again consistent with the experiment (Fig. 4 in Adamo et al. (2010)).

The Cue N2pc and Target N2pc components observed in the experiment can be explained by the dynamics of the presented model. The V4pool layer of the model receives reentrant signals of spatial attention from FEFvm layer. These spatial reentrant signals focus the attention over one location and induce a competition between locations. After the initial rise elicited by the cues, a further increase is caused by reentrant processing via the FEFvm, visible in the Cue N2pc signal (Fig. 12). Reentrant processing is also visible in the mean FEFvm activities (Fig. 14). The FEFvm activities are presented for the case that the target stimulus is blue on left. While the S+C+ and S-C- conditions are quite obvious, a comparison between S+C- and S-C+ is particularly interesting. In the S-C+ condition, the cue with the target color is at the wrong location and thus does not provide a strong reentrant processing of the target, while the cue in the S+C- condition, although it does not match the full ACS, it matches the color target template and thus leads to a strong spatial reentrant processing at the target location.

4. Discussion

The phenomenon of how our attention system processes two distinct visual attentional control sets at a time has been debated much recently, but has not been investigated by neuro-computational models. We designed and simulated a neuro-computational model of attention and decision making to better understand the neural correlates of how two ACSs capture attention. We simulated the experiment of Adamo et al. (2010) who asked subjects to focus their attention on two different colors combined with two different locations. Cues presented prior to the target influence the decision time of the network towards the target stimulus. Importantly, in the experiment of Adamo et al. (2010), all locations are cued by changing the luminance, so that the match to the ACS is relevant (Anderson & Folk, 2012), but not any stimulus-driven attention (Belopolsky et al., 2010). Therefore, consistent with contingent capture theory (Folk et al., 1992), the simulation results can be interpreted based on the effect of the top-down effects of ACSs and the congruency of cue and target. However, given the hypothesis that attention is initially driven by a global color-based signal, the observations of Adamo et al. (2010) are not easy to interpret. The reaction time pattern has been interpreted by Adamo et al. (2008) as an indication for two ACS for color and location. Our model suggests that the reaction time data is distorted by decision priming, as we could only replicate the behavioral data when we start to integrate already upon cue information. Further, our model indicates some additional priming, such as object/feature and location priming. Inter trial priming reported in some studies (Büsel, Pomper, & Ansorge, 2019; Irons et al., 2012) may also play some role, but our model does not include any properties for this.

Importantly, the model also replicates the electrophysiological data of Adamo et al. (2010), which allows us to take it as a basis to better understand the involved neural processes. Visualizing the dynamics of neural activities in the model enables us to understand the observations of Adamo et al. (2010) about the congruency of cue and target stimulus and also the congruency of cue and attentional control sets. Adamo et al. (2010) found that the N2pc to cues that fit an ACS, called good cues, does not significantly differ from the N2pc to the one of cues that do not fit to the ACS, called bad cues, and conclude that all cues capture

attention. This may be surprising as the nature of the cue has a clear effect on the reaction time of the subjects. Our model suggests that the cue image is not filtered early by its match to the ACS to determine the amount of spatial attention, but that initially, only a feature-based attention signal for both target colors operates in parallel, regardless of the location of stimuli in the scene. This model assumption is consistent with the conclusions derived by Becker, Ravizza, and Peltier (2015), Berggren, Jenkins, McCants, and Eimer (2017), and Irons and Remington (2013) who proposed that colors of two ACSs are attended globally. This explains that cues, which have a full match to the ACS, do not initially capture more attention than those, which only have a color match.

This early feature-based bias also explains the observation by Adamo et al. (2010) that the Target N2pc latency is shorter for congruent location (S+), as here, a cue that matches the ACS in color is at the same location as the target. Any cue that is consistent with the feature-based attention signal gets slightly more enhanced than any other cue. This slightly more enhanced signal contributes to the accumulation process, but in addition, is sent to the FEF and re-enters the visual cortex and enhances the response to the target. Such combination of feature and re-entrant spatial attention is a core component of our model (Hamker, 2003; Hamker, 2005; Moore & Armstrong, 2003) and helps to better understand the complex nature of the data from Adamo et al. (2010). For example, the models like Bundesen, Habekost, and Kyllingsbæk (2005) may not explain this observation. Because first, in their model, when the model is cued by location, the pertinence values which determine the attentional weights, are determined before stimulus presentation and when the model is cued by feature, the pertinence values are determined longer after stimulus presentation. Therefore, they do not apply the spatial feedback in their model. Second, the selection mechanism in their model is based on identifying the category of objects rather than spatial selection. Denison et al. (2021) presented a model for the dynamics of temporal attention using a normalization model implemented in three layers of sensory, attention and decision making, However, their model does not include a spatial reentrant processing loop, which is crucial for attentional capture. Similarly, Guided Search 6.0 (Wolfe, 2021) mainly aims at replicating set-size dependent reaction times, but it has no natural model dynamics to provide a component comparable to N2pc. It assumes a very fast scanning of attention (20 times per second), which is different from our model. A following binding and recognition process modelled by drift diffusion is similar to our decision process.

With respect to Target P3, Adamo et al. (2010) found differential effects of P3 amplitude and P3 latency. P3 amplitude mainly depended on whether the cue has the same color as the target but not if the cue fully matches the ACS. P3 latency is particularly interesting for cues presented at the same side as the target, as latency is shortest if the cue matches the color of the ACS (S+C+) and longest, if they do not match the ACS (S+C-). Although the Target P3 includes processes somewhat beyond what we can simulate with our model, we can at least provide some intuition. P3 amplitude could be affected by the recurrent loop between V4gain and V4pool cells, e.g. recurrent processing in the visual cortex that boosts features that match between the preceding cue and the target, in this experiment their color. The short P3 latency is likely affected by the re-entrant processing via the FEF, as a cue that does not match the ACS (C-) leads to weaker activation of the FEF neurons which in turn provides a weaker gain to the target leading to a reduced loop activity.

When we relate physiological measures such as N2pc to the amount of reentry from the FEF in our model, N2pc measures the relative difference of attention between two locations, while the reentry from the FEF in our model (but see also Juan, Shorter-Jacobi, & Schall, 2004; Ray, Pouget, & Schall, 2009) indicates the overall amount of attention directed to a particular location. This interpretation is consistent with previous experimental (Dubois et al., 2009) and computational (Zirnsak et al., 2011) studies, which suggested that transiently, attention can be directed to multiple non-contingent items until it converges into a sustained mode, where the focus of attention is on one item.

To summarize, the results of our computational model support the idea that subjects do not filter the visual scene initially by feature and color, but apply a global feature search. Reaction time data that suggests ACS for different locations in space, e.g. Adamo et al. (2010), Adamo et al. (2008), could be distorted by priming as predicted by our model.

Credit authorship contribution statement

Shabnam Novin: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Visualization,**Ali Fallah:**Supervision, **Saeid Rashidi:**Project administration,**Frederik Beuth:**Methodology, Writing – review & editing,**Fred H. Hamker:**Supervision, Conceptualization, Methodology, Writing – review & editing, Funding acquisition.

Acknowledgments

This research has been supported by the grant of Cognitive Sciences and Technologies Council of Iran, European Social Fund (ESF), and DAAD Scholarship STIBET III and in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 416228727 – SFB 1410 Hybrid Societies.

References

Adamo, M., Pun, C., & Ferber, S. (2010). Multiple attentional control settings influence late attentional selection but do not provide an early attentional filter. *Cognitive Neuroscience, 1*(2), 102–110.

Adamo, M., Pun, C., Pratt, J., & Ferber, S. (2008). Your divided attention, please! The maintenance of multiple attentional control sets over distinct regions in space. *Cognition, 107*(1), 295–303.

Andersen, S. K., Fuchs, S., & Müller, M. M. (2011). Effects of feature-selective and spatial attention at different stages of visual processing. *Journal of Cognitive Neuroscience, 23*(1), 238–246.

Anderson, B. A., & Folk, C. L. (2012). Dissociating location-specific inhibition and attention shifts: Evidence against the disengagement account of contingent capture. *Attention, Perception, & Psychophysics, 74*(6), 1183–1198.

Awh, E., & Pashler, H. (2000). Evidence for Split Attentional Foci. *Journal of Experimental Psychology, 26*(2), 834–846.

Becker, M. W., Ravizza, S. M., & Peltier, C. (2015). An inability to set independent attentional control settings by hemifield. *Attention, Perception, & Psychophysics, 77*(8), 2640–2652.

Belopolsky, A. V., Schreij, D., & Theeuwes, J. (2010). What is top-down about contingent capture? *Attention, Perception, & Psychophysics, 72*(2), 326–341.

Berggren, N., Jenkins, M., McCants, C. W., & Eimer, M. (2017). The spatially global control of attentional target selection in visual search. *Visual Cognition, 25*(1-3), 196–214.

Beuth, F. (2019). *Visual attention in primates and for machines - neuronal mechanisms.* Doctoral Thesis (pp. 1–284). Germany: Technische Universität Chemnitz.

Beuth, F., & Hamker, F. H. (2015). A mechanistic cortical microcircuit of attention for amplification, normalization and suppression. *Vision Research, 116*, 241–257.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A Neural Theory of Visual Attention: Bridging Cognition and Neurophysiology. *Psychological Review, 112*(2), 291–328.

Büsel, C., Pomper, U., & Ansorge, U. (2019). Capture of attention by target-similar cues during dual-color search reflects reactive control among top-down selected attentional control settings. *Psychonomic Bulletin & Review, 26*(2), 531–537.

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525.

Cave, K. R., Bush, W. S., & Taylor, T. G. G. (2010). Postscript: Two Separate Questions in Split Attention: Capacity for Recognition and Flexibility of Attentional Control. *Psychological Review, 117*(2), 695–696.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the stroop effect. *Psychological Review, 97*(3), 332–361.

Cohen, J. Y., Pouget, P., Heitz, R. P., Woodman, G. F., & Schall, J. D. (2009). Biophysical support for functionally distinct cell types in the frontal eye field. *Journal of Neurophysiology, 101*(2), 912–916.

Dannhauser, T. M., Walker, Z., Stevens, T., Lee, L., Seal, M., & Shergill, S. S. (2005). The functional anatomy of divided attention in amnestic mild cognitive impairment. *Brain, 128*(6), 1418–1427.

Denison, R. N., Carrasco, M., & Heeger, D. J. (2021). A dynamic normalization model of temporal attention. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01129-1

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

Dubois, J., Hamker, F. H., & VanRullen, R. (2009). Attentional selection of noncontiguous locations: The spotlight is only transiently "split". *Journal of Vision, 9*(5:3), 1–11.
Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in Cognitive Sciences, 18*(10), 526–535.
Folk, C., Remington, R., & Johnston, J. (1992). Involuntary Covert Orienting Is Contingent on Attentional Control Settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 1030–1044.
Frey, H. P., Schmid, A. M., Murphy, J. W., Molholm, S., Lalor, E. C., & Foxe, J. J. (2014). Modulation of early cortical processing during divided attention to non-contiguous locations. *European Journal of Neuroscience, 39*(9), 1499–1507.
Gaillard, C., Hassen, S. B. H., Di Bello, F., Bihan-Poudec, Y., VanRullen, R., & Hamed, S. B. (2020). Prefrontal attentional saccades explore space rhythmically. *Nature Communications, 11*(1), 1–13.
Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience, 4*(7), 563–572.
Grubert, A., & Eimer, M. (2016). All set, indeed! N2pc components reveal simultaneous attentional control settings for multiple target colors. *Journal of Experimental Psychology. Human Perception and Performance, 42*(8), 1215–1230.
Hamker, F. H. (2003). The reentry hypothesis: Linking eye movements to visual perception. *Journal of Vision, 3*(11), 808–816.
Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research, 44*(5), 501–521.
Hamker, F. H. (2005). The reentry hypothesis: The putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral Cortex, 15*(4), 431–447.
Heinze, H. J., Luck, S. J., Munte, T. F., Gös, A., Mangun, G. R., & Hillyard, S. A. (1994). Attention to adjacent and separate positions in space: An electrophysiological analysis. *Perception & Psychophysics, 56*(1), 42–52.
Hüttermann, S., & Memmert, D. (2017). The attention window: A narrative review of limitations and opportunities influencing the focus of attention. *Research Quarterly for Exercise and Sport, 88*(2), 169–183.
Irons, J. L., Folk, C. L., & Remington, R. W. (2012). All set! Evidence of simultaneous attentional control settings for multiple target colors. *Journal of Experimental Psychology: Human Perception and Performance, 38*(3), 758–775.
Irons, J. L., & Remington, R. W. (2013). Can attentional control settings be maintained for two color–location conjunctions? Evidence from an RSVP task. *Attention, Perception, & Psychophysics, 75*(5), 862–875.
Jans, B., Peters, J. C., & De Weerd, P. (2010). Visual spatial attention to multiple locations at once: The jury is still out. *Psychological Review, 117*(2), 637–682.
Jia, J., Liu, L., Fang, F., & Luo, H. (2017). Sequential sampling of visual objects during sustained attention. *PLoS Biology, 15*(6), e2001903.
Juan, C. H., Shorter-Jacobi, S. M., & Schall, J. D. (2004). Dissociation of spatial attention and saccade preparation. *Proceedings of the National Academy of Sciences, 101*(43), 15541–15544.
Kawahara, J. I., & Kumada, T. (2017). Multiple attentional sets while monitoring rapid serial visual presentations. *Quarterly Journal of Experimental Psychology, 70*(11), 2271–2289.
LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96*(1), 101–124.
Liu, T., & Jigo, M. (2017). Limits in feature-based attention to multiple colors. *Attention, Perception, & Psychophysics, 79*(8), 2327–2337.
Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature, 421*(6921), 370–373.
Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*(2), 160–174.
Ray, S., Pouget, P., & Schall, J. D. (2009). Functional distinction between visuomovement and movement neurons in macaque frontal eye field during saccade countermanding. *Journal of Neurophysiology, 102*(6), 3091–3100.
Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5*(7), 631–632.
Schall, J. D. (1991). Neuronal activity related to visually guided saccades in the frontal eye fields of rhesus monkeys: Comparison with supplementary eye fields. *J Neurophysiol, 66*(2), 559–579.
Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron, 55*(2), 301–312.
Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, 1–33.
Zirnsak, M., Beuth, F., & Hamker, F. H. (2011). Split of spatial attention as predicted by a systems-level model of visual attention. *European Journal of Neuroscience, 33*(11), 2035–2045.