Neuropsychologia 47 (2009) 2354–2366

Contents lists available at ScienceDirect

Neuropsychologia

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

# Modelling distractor devaluation (DD) and its neurophysiological correlates

Nickolaos Fragopanagos[a,*], Tamara Cristescu[b], Brian A. Goolsby[c], Monika Kiss[d], Martin Eimer[d], Anna C. Nobre[b], Jane E. Raymond[c], Kimron L. Shapiro[c], John G. Taylor[a]

[a] King's College London, UK
[b] University of Oxford, UK
[c] University of Wales, Bangor, UK
[d] Birkbeck College, University of London, United Kingdom

ARTICLE INFO

*Article history:*
Received 7 September 2007
Received in revised form
10 September 2008
Accepted 14 September 2008
Available online 27 September 2008

*Keywords:*
Attention
Emotion
Simulation
N2pc
Faces

ABSTRACT

A series of recent studies have shown that selective attention can influence the emotional value of both *selected* as well as *ignored* items. Specifically, ignored items (distractors) were consistently rated less positively in emotional evaluations, following attentional selection, relative to (typically) simultaneously presented items (targets). Furthermore, a known electrophysiological index of attentional selectivity (N2pc) was shown to correlate with the magnitude of the observed 'distractor devaluation' (DD). A neural model is presented here to account for these findings by means of a plausible mechanism linking attentional processes to emotional evaluations. This mechanism relies on the transformation of attentional inhibition of the distractor into a reduction of the value of that distractor. The model is successful in reproducing the existent behavioural results as well as the observed link between the magnitude of the attentional N2pc and the magnitude of DD. Moreover, the model proposes a series of testable hypotheses as well as specific predictions that call for further experimental investigation.



## 1. Introduction

Attention and emotion are neural systems involved in prioritisation across available objects for response, the former by facilitating their perceptual processing and the latter by assessing their value, both, according to current and future goals. Numerous studies have shown that emotion can strongly direct the focus of attention (Eastwood, Smilek, & Merikle, 2001; Fenske & Eastwood, 2003; Fox, Russo, Bowles, & Dutton, 2001). Attention can also be captured by sensory items of high emotional value, such as emotional faces (Vuilleumier & Schwartz, 2001). We have presented elsewhere our attention/emotion model, which is able to reproduce the findings from a variety of such studies (Taylor & Fragopanagos, 2005). However, a major question remains: can the reciprocal process occur, that is, can attentional selection have an influence on the emotional value of the items attention acts upon, namely the target and the distractor(s)? A series of recent studies by Raymond and colleagues suggest an affirmative answer (Fenske, Raymond, Kessler, Westoby, & Tipper, 2005; Fenske, Raymond, & Kunar, 2004; Raymond, Fenske, & Tavassoli, 2003; Raymond, Fenske, & Westoby, 2005). These investigators used a simple visual search paradigm where participants were required to make a feature-based selection between two-coloured abstract images, one containing the target feature and one not. To investigate the existence of attentional effects on emotional evaluation, they asked participants following the attentional selection to rate the target, the distractor, or a novel image for their emotional value using a cheerful/dreary dimension. Their results confirmed the hypothesis showing that distractor images were rated as less cheerful (or more dreary) than target or novel images, which did not differ from each other. Given that each unique pattern was unfamiliar and seen on only a single trial, only the prior attentional state can explain this modulation in evaluative judgments. Also, the effect seems to be driven by ignoring of the distractor items, since distractors were devalued relative to targets and novel items. Raymond and colleagues termed this difference in evaluative ratings of prior targets and distractors the "distractor devaluation" (DD) effect and suggested that it resulted from a re-instantiation of the inhibition applied to the distractor during selection. Using a more complex multi-item visual search paradigm, Fenske et al. (2004) found that ratings for previewed distractors were lower than distractors onsetting with the target, consistent with the initial account of visual marking

*Abbreviations:* DD, distractor devaluation; ME, mere exposure; PFC, prefrontal cortex; PPC, posterior parietal cortex; FFA, fusiform face area; OFC, orbitofrontal cortex; AMG, amygdala.

* Corresponding author at: Mathematics Department, King's College London, Strand, London WC2R 2LS, UK. Fax: +44 2078482017.
*E-mail address:* nickolaos.fragopanagos@kcl.ac.uk (N. Fragopanagos).

0028-3932/$ – see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.neuropsychologia.2008.09.018

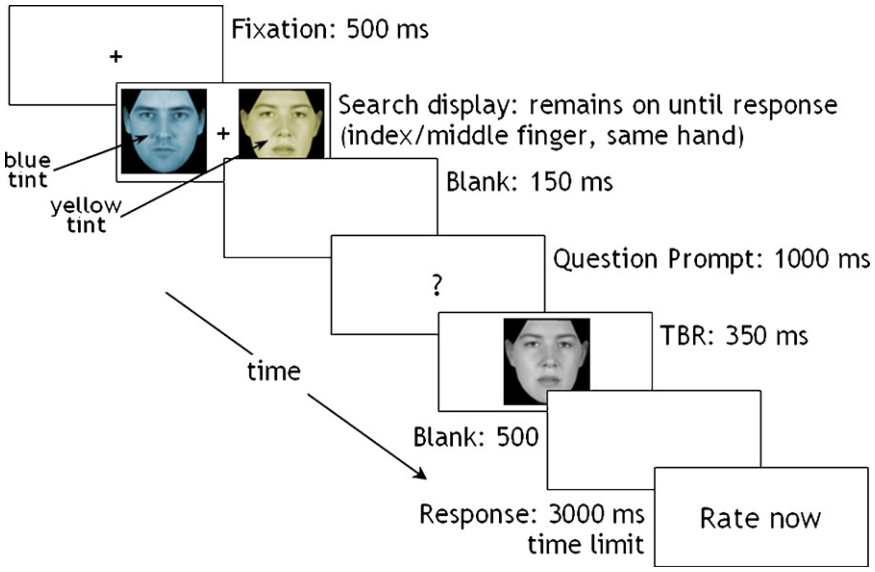


**Fig. 1.** The sequence of stimuli in an example trial from the Goolsby et al. (2007) study. The trial began with the attention task followed by the evaluation task. A two-face display followed a 1000 ms presentation of a fixation cross. In experiments 2 and 3, the two-face display remained on the screen until observers made a response whereas in experiment 3 the two-face display was restricted to 200 ms. After selecting the target and reporting the relevant feature, a blank screen appeared for 150 ms, followed by a "?" prompt appearing for 1000 ms preparing participants for the trustworthiness evaluations. Then, a greyscale version of the face for evaluation was presented at the centre of the screen for 350 ms followed by a blank screen appearing for 500 ms. Finally, the prompt "rate now" appeared and remained on-screen until the rating response or 3000 ms, whichever came first.

(Watson & Humphreys, 1997) arising from inhibition of the previewed distractors. Raymond et al. (2005) investigated the effect of target–distractor spatial distance on the magnitude of the DD and found DD to be larger when the distractors were closer to the targets' location. This was also the first study (experiment 3) to report a robust DD effect using neutral-expression faces that were later rated on a social-emotional dimension (trustworthiness). Finally, Fenske et al. (2005), also using neutral faces and trustworthiness judgements in a simple go/no-go task, showed that faces associated with a no-go cue were judged as less trustworthy than uncued faces, suggesting that response inhibition could also modulate social-emotional judgements.

We now turn to two recent studies whose findings form the basis for the DD model presented in the next section. The first study was carried out by Goolsby et al. (2009) in order to investigate the boundary conditions of the DD effect. The basic paradigm used in this study (henceforth referred to as the 'two-face' paradigm) involved an attentional selection phase where two neutral faces were shown bilaterally to fixation; participants were required to locate the target (defined by a feature along which the two faces differed, e.g., gender) and report another feature of the target, e.g., tint colour. Subsequent to this selection, participants were asked to rate one of the two faces shown at the selection phase or a novel face on a trustworthiness scale similar to that used in Raymond et al. (2005). Fig. 1 illustrates a sample trial.

The results from this study showed that when the feature on which the attentional selection was based was not present for the trustworthiness evaluation, no DD effect was observed. This was the case in experiment 1 where the selection feature (colour tint) was removed at evaluation (greyscale face presentation). On the other hand, a robust DD was obtained when the attentional selection feature was present during evaluation. This was observed in experiments 2 and 3 where the selection feature was the gender of the faces which was obviously present at evaluation. This finding indicates that the DD effect is feature-based and not exemplar-based in that the identity of the distractor face was shown not to be sufficient for DD to occur. Further support for this conclusion is given by a careful analysis of the behavioural results conducted by the authors and reproduced here in Fig. 2. In this figure the average trustworthiness evaluations are plotted separately for the targets, the distractors, target-like novels and distractor-like novels. In the case of experiments 2 and 3 (plotted separately in Fig. 2), where the selection feature was the gender, target-like novels were novel faces of the target gender whereas distractor-like novels were novel faces of the distractor gender. From this figure one can clearly see that DD occurs for the distractor-like novels as well as for the actual distractors although the former were not shown prior to their trustworthiness evaluations and, therefore, were not explicitly affected by any attentional processing. It appears that presentation of even a new stimulus for trustworthiness evaluation is enough to obtain a DD effect as long as it bears the distractor-defining feature. An obvious difference apparent in Fig. 2 is that the actual targets and distractors appear to be judged slightly more trustworthy compared to their novel counterparts (although this difference is non-significant). We have attributed this difference

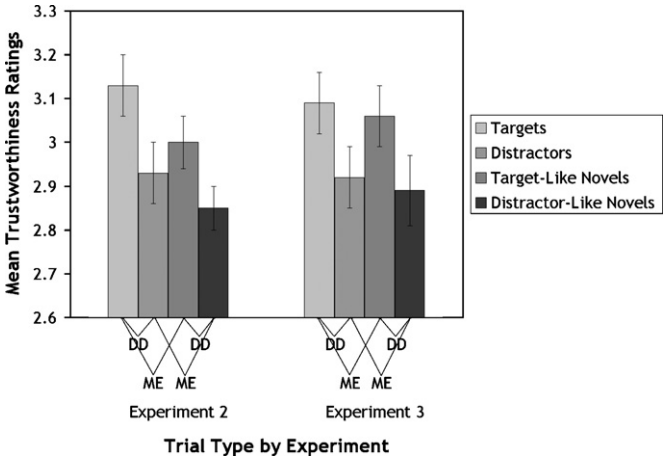


**Fig. 2.** Behavioural results from Goolsby et al. (2007) suggesting a mixture of distractor devaluation (DD) and mere exposure (ME) underlie the effect.

to the mere exposure (ME) effect, which predicts a positive affective boost when stimuli are repeatedly presented (Kunst-Wilson & Zajonc, 1980). Thus, the resulting behavioural effect in Goolsby et al. (2009), we propose, is a mixture of DD and ME that can be deconstructed in plots such as those in Fig. 2. This interpretation forms the basis of our computational model that is presented in detail in the following section.

The two-face paradigm of Goolsby et al. (2009) was also used in an electrophysiological study (Kiss et al., 2007) that aimed to delineate the neural mechanism underlying the DD effect as well as to provide an additional replication of the behavioural results of Goolsby et al. (2009). The authors focused on a specific component of the event-related potentials (ERPs) known as the N2pc. The N2pc is defined as the excess of negative voltage over posterior scalp sites contralateral to the target relative to the ipsilateral ones. These voltages are recorded from posterior lateralised electrodes (usually the occipito-parietal PO7 and PO8) and the excess negativity typically occurs roughly in the N2 latency range (180–300 ms). The N2pc has been shown in numerous studies to correlate with target selection and/or distractor inhibition (Eimer, 1996; Luck & Hillyard, 1994; Woodman & Luck, 1999), and provides a reliable measure of attentional processing/capture. Motivated by the attentional inhibition account of the DD effect by Raymond and colleagues (Fenske et al., 2004, 2005; Goolsby et al., 2009; Raymond et al., 2003, 2005), the study of Kiss et al. (2007) aimed to test the hypothesis that more efficient attentional target selection, which might be mediated by more efficient inhibition of distractor items, as reflected by larger N2pc amplitudes, could be related to more devaluation of distractors. The paradigm used in this study was almost identical to that of experiment 3 in Goolsby et al. (2009). Two faces were shown on either side of a fixation cross for 200 ms, and the task was to select the face with the target gender and report its colour. After selection and a fixation interval of 1200 ms, one of the two faces was shown again centrally for 350 ms followed by a blank interval of 500 ms after which participants had to evaluate the single face on a trustworthiness scale. The EEG signal was time-locked to the onset of the search array, and the N2pc measured during attentional selection was examined as a function of prior trial type (target or distractor) and trustworthiness rating (high or low) during the evaluation phase. Indeed, confirming the initial hypothesis, Kiss et al. (2007) found that when distractors were rated low, the N2pc was large (suggesting efficient selection), but when distractors were rated high, the N2pc was small and delayed (suggesting a more diffuse spread of attention during selection). However, the same link was not observed for the evaluation of the target stimuli, for which the amplitude of the N2pc was unrelated to the subsequent targets' evaluations. This set of results points to an inhibition-based account for the DD effect, whereby the more inhibition applied during attentional selection (or the bigger the N2pc), the more distractor devaluation ensues. In the following sections we will demonstrate how we incorporate this account in our model and reproduce the results from the Kiss et al. (2007) study.

## 2. Methods

The model we have constructed in order to simulate the DD effect consists roughly of two parts: one is a network of occipital/temporal and frontoparietal modules that deal with the perceptual processing of the visual stimuli and the allocation of attention to them, respectively, and the other is a network of limbic modules that deal with the encoding of the internal value of the stimuli. The former network can be considered to play the attentional role in the model and the latter the emotional. We have developed similar cognitive/emotional models in the past in order to computationally investigate the effect of emotional load on the perception/attention system (Taylor & Fragopanagos, 2005). As stated in Section 1, presently we were interested in modelling the reciprocal effect, that is, the effect of attentional load on emotional value. For that reason, in linking the two parts of the model, we focused on the connections from the perception/attention system to the emotional. The overall architecture of the model is based on substantial volume of research on the perceptual and attention systems and a reasonable amount of research on the emotion system. The links between the perception/attention system and the emotion system in the model are also broadly based on the known anatomical connectivity from both primate and human neuroanatomy. However, we have made a number of assumptions regarding the more detailed characteristics of these links that were motivated mainly by the extant literature on the DD effect, itself. As such, the model goes beyond a mere reproduction of the extant results from the DD effect studies by proposing a range of hypotheses that can be tested experimentally both behaviourally and neuroscientifically. These hypotheses will be discussed after the presentation of the current results below. Before beginning the more detailed description of the various components of our model we present a diagram showing the overall architecture of the model and connections between the modules.

Let us turn to the building blocks of the model in more detail. The entry point for external stimuli in our model is the visual system. This, in humans, is known to comprise a wealth of brain structures mostly located in and around the occipital and the temporal lobes. These areas are usually defined functionally and, in this regard, they exhibit a certain degree of hierarchical organisation. The lower-tier areas deal with the simple features of visual objects and are highly topographical while the higher-tier areas deal with complex features and their conjunctions (or even whole objects) and retain very coarse retinotopy (if any). In the paradigm that we are trying to simulate, participants are asked to find the face (between two candidates) that is of the target gender (and report its colour). This, in other words, is a face-based gender discrimination problem, the solution of which in our model relies on a simplified ventral-route visual system. Furthermore, as it is difficult to identify the contribution of the lower-tier visual areas to the gender discrimination task, we have only included in our model a V4 module (known to process complex features and, thus, suitable for processing gender-related features) and a fusiform face area (FFA) module known to be particularly sensitive to faces and, thus, also suitable to contribute to the task. Neurons in the superior temporal sulcus (STS) have also been found to respond preferentially to faces (especially in monkeys, Tsao, Freiwald, Tootell, & Livingstone, 2006), but an STS module was not included in the model at this time as the task could be handled without it.

Since the paradigm involves two faces presented bilaterally to a central fixation cross, our V4 module consists of a left-hemisphere and a right-hemisphere component that each process the contralateral face (ipsilateral activations can be neglected for simplicity). Each hemispheric component contains three neurons, one coding for male features, one for female features and one for non-gender-specific or ambiguous features. In Fig. 4 (depicting the model's architecture) the latter is represented by a superposition of the male and female biological symbols whereas the male-feature-coding neurons are represented by the male biological symbol and the female-feature-coding neurons by the female one. It is clear that our model's single neurons represent thousands or tens of thousands of real brain neurons and thus are abstractions. It is important to note here that there is lateral inhibition between the male-feature-coding and the female-feature-coding neurons in each hemisphere. The FFA module consists of two neurons, one coding for male faces and one coding for female faces. We have taken these neurons' receptive fields to be large enough to cover the entire visual display and, thus, each neuron can be activated by either left- or right-presented face. These neurons are also represented in the model's diagram (Fig. 4) by the respective biological symbols and are also abstractions. Each FFA neuron coding for a specific gender is receiving its input from the neurons in the V4 module that code for that gender's features as well as from the non-gender-specific (androgynous) neurons. Crucially, each FFA neuron also feeds back to the same-gender V4 neurons, a particularly important feature that mediates attentional control as described below.

We now turn to the network of modules involved in attentional control. The bilateral face paradigm engages the attention system under two conditions: one is endogenous as defined by the task goal (find male/female face) and is feature-based and the other is exogenous as a reflex to the consequences of the endogenous attention condition (male/female face is on left/right) and is of a spatial nature. Endogenous and exogenous attention are known to be subserved by a largely overlapping network of frontoparietal areas (Kim et al., 1999; Nobre et al., 1997) although some studies have found a certain degree of specialisation for endogenous or exogenous attention within that network (Corbetta & Shulman, 2002; Kincade, Abrams, Astafiev, Shulman, & Corbetta, 2005; Yantis et al., 2002). Similarly, spatial attention and feature-based attention have been shown to engage common frontoparietal areas (Arrington, Carr, Mayer, & Rao, 2000; O'Craven, Downing, & Kanwisher, 1999; Wojciulik and Kanwisher, 1999) with spatial attention possibly activating dorsal areas more strongly than feature-based attention (Giesbrecht, Woldorff, Song, & Mangun, 2003). In our model, for simplicity, we have designated a prefrontal cortex (PFC) module to the endogenous feature-based attentional control and a posterior parietal cortex (PPC) module to the exogenous spatial attention. The PFC module is very similar to the FFA module as it contains the two neurons representing the male and the female face. These two neurons are not activated by external stimulation but rather by the task goal (select male/female). The PFC neurons are connected excitatorily to their counterparts in the FFA module and inhibitorily to the opposite-gender FFA neurons. The latter is actually achieved by excitatory connections to local inhibitory interneurons so that the overall effect is inhibitory. So, one of the two PFC neurons is activated according to the task instructions (find male/female

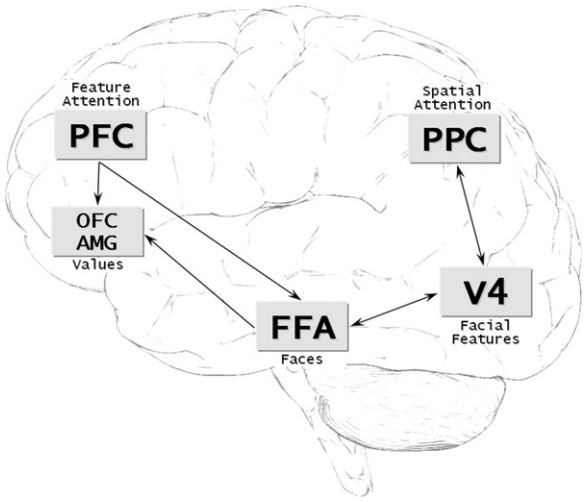


**Fig. 3.** Module-level architecture of DD model. Each box corresponds to one of the model's modules with its suggested brain area siting indicated in the box and the module function printed below or above the box. Further details follow in main text below.

face) and then feeds back activity to the same-gender FFA neuron to favourably bias it in preparation for the onset of the face stimuli. The same PFC neuron also feeds back inhibitorily to the opposite-gender neuron as described above. Due to the connections between the FFA and the V4 neurons, this PFC-driven attention signal cascades back to the V4 neurons such that they also exhibit a favourable bias of the target-gender feature neurons (in both left and right hemispheric components). The PPC module comprises two neurons, one sensitive to left-hemisphere visual stimulation and the other for the right-hemisphere. However, being part of the dorsal 'where' circuit, these neurons cannot discriminate object features (unlike their ventral 'what' counterparts). Similar to the PFC neurons, the PPC neurons have a same-hemisphere excitatory and opposite-hemisphere inhibitory connectivity with the V4 neurons. More specifically, the left-hemisphere PPC neuron connects excitatorily to all the left V4 neurons and inhibitorily to all the right V4 neurons; the converse is true for the right-hemisphere PPC neuron. This way, the PPC module acts as a spatial attention controller that is able to selectively bias activations in one hemisphere depending on where the salient object is being processed. This mechanism is consistent with a recent study of feature-based attention where it was shown that, during visual search, feature-based attention precedes the spatial allocation of attention as the latter is only engaged once sufficient target discrimination has been achieved (Hopf, Boelmans, Schoenfeld, Luck, & Heinze, 2004). We do not wish to specify the precise location of the PFC and the PPC neurons within the PFC and the PPC, respectively. Our intention, on the other hand, was to emphasise the endogenous character of the feature-based attentional selection as opposed to the more reflexive spatial attention that was assigned to the PPC module.

The general principle underlying the attentional network in our model is Desimone and Duncan's 'biased-competition hypothesis' (Desimone & Duncan, 1995), whereby, when multiple stimuli compete for representation at a certain location, attention can be used to bias this competition favourably for the behaviourally relevant stimulus, which, thereby, eventually wins the competition. Several computational models of attention have successfully implemented the concept of 'biased-competition' in order to account for a range of experimental findings of attention studies (Deco & Rolls, 2003; Deco & Zihl, 2001; Hamker, 2004; Usher & Niebur, 1996). Furthermore, many computational models assume that the focus of attention is moving around in a spatial map until it finds the most 'salient' position on that map. The computation of the most 'salient' position on the map can be performed in a purely bottom-up fashion (Itti & Koch, 2000) or in relation to the task imperatives, such as in Wolfe's Guided Search Model (Wolfe, 1994). The latter uses top-down feature maps in order to guide the visual search for the target. However the construction of these top-down feature maps in this and other, related, models is mechanistic and ad hoc. Tsotsos et al. (1995) have developed a model that uses a combination of strong WTA competition within each layer of the ventral hierarchy and top-down inhibition as a means to form a spatially focused attentional spotlight. This is an interesting approach to solving the top-down/bottom-up integration problem; however, it ties the feature-based top-down modulation too tightly to the spatial focus of attention. Our model follows very closely that of Hamker (2004) where top-down and bottom-up information is integrated in the V4 layer with feature-based modulation of V4 nodes leading to the highlighting of a location in perceptual and premotor maps which, in turn, feedback spatially to further enhance target selection in V4.

The model we propose in Figs. 3 and 4 is thus based on general well-accepted principles of attention as a feedback control system with components based as posited in Fig. 3 and with detailed neural structure as in Fig. 4. The components have been chosen to be as simple as possible, both in number and in internal structure, so as to demonstrate as transparently as possible the manner in which the processing takes place. In particular we are interested in the interaction of the top-down and bottom-up attention controls for a subject to achieve the required target face and avoid the distracter (leading thereby to the N2pc), as well as how this is translated into devaluation in the limbic system.

We can embed the model of Figs. 3 and 4 into the more general CODAM model of attention (Fragopanagos, Kockelkoren, & Taylor, 2005; Taylor, 2003), which treats attention as a control signal whose purpose is to amplify the lower level cortical input activity representing an external stimulus to attain activation on a suitable working memory site. The CODAM model was based on engineering control theory. The CODAM model allows for the presence of a range of component modules functioning as possible internal models and predictors, as well as error monitors. Thus we can regard the simulation development as a part of the program of testing the CODAM model in the case of the interaction of top-down and bottom-up attention control.

Let us now turn to the emotional part of our model and see how it is linked to the perception/attention part described above. The emotional part of our model has been collapsed into a single module, which represents the orbitofrontal cortex (OFC) and the amygdala (AMG). The OFC has been found to contain populations of neurons that encode abstract reward and punishment in humans using fMRI (O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001) as well as in primates using single cell recordings (Hosokawa, Kato, Inoue, & Mikami, 2007). The primate amygdala has also been shown to represent positive and negative value of visual stimuli during learning (Paton, Belova, Morrison, & Salzman, 2006). Moreover, both OFC and AMG are known to be well connected to regions in temporal cortex coding for visual objects (Cavada, Company, Tejedor, Cruz-Rizzolo, & Reinoso-Suarez, 2000; Freese & Amaral, 2005). Based on these results we have included two pairs of neurons in the OFC/AMG module, one pair of reward and punishment neurons for the male face and another pair of reward and punishment neurons for the female face. The reason we have used only neurons coding for the value of a face feature (the gender) and not neurons coding for value of specific exemplar faces is the fact that the DD effect does not appear to depend on the specific faces shown during the attention stage of the paradigm (Goolsby et al., 2009); rather it seems to operate at a featural level as suggested by its generalisation to novel faces as discussed in Section 1. The OFC/AMG neurons are activated by the same-gender FFA neurons whenever faces are presented, thus providing a neural link between the faces and their values. Crucially, the OFC/AMG neurons also receive input (during the attentional selection phase) from the PFC neurons. These inputs from the PFC are essentially axon collaterals of the long PFC axons that reach the inhibitory interneurons in the FFA to cause distractor inhibition there. In this way, the distractor inhibition (attention-related) is "copied" onto the OFC/AMG module as an increase of the distractor punishment neuronal activation (emotion-related). This change of the distractor punishment neuronal activation causes the observed DD, which occurs at the trustworthiness evaluation stage; thus there must be some persistent change (which nevertheless decays with time) that carries over from the attentional stage into the trustworthiness evaluation stage. We have employed a modified Hebbian learning of the FFA to OFC/AMG weights that includes a leakage term forcing the weight adaptation to decay when not reinforced. Thus the equation for the updating of the DD-related weights is:

$$\tau \frac{dw}{dt} = -w + (\text{input})(\text{output}).$$

As mentioned in Section 1, the behavioural results from the bilateral face paradigm indicate that faces that have been presented during the attentional selection stage (as either target or distractor), receive a small amount of positive bias in the trustworthiness evaluation stage relevant to faces not shown prior (novel faces). We have attributed this difference to an ME contribution. Thus we had to include a mechanism in our model that can account for this contribution. There are several theories that attempt to explain how ME works. These theories tend to be of a theoretical form and have not yet been made neurally plausible. The most prominent account of ME is the perceptual fluency theory (Bornstein & D'Agostino, 1994; Seamon, Brody, & Kauff, 1983) whereby repeated stimulus exposure leads to greater processing fluency which is misattributed as greater preference on subsequent encounters. Although this theory proposes that exposure has affective consequences, these occur only at the time of evaluation (as a misattribution of cognitive fluency) and not during the initial exposure. However there is some recent evidence (Harmon-Jones & Allen, 2001) which suggests that affect is involved in the ME effect (as indexed by increased zygomatic activity and reduced left frontal cortical activation). We have thus adopted a purely limbic account for ME for our model (although a mixture of affective and more cognitive components are more likely to comprise the basis of ME). This approach uses two extra nodes in the OFC/AMG module for the prior target and distractor that are being fed by an equivalent pair in the FFA module. The connection weights between the FFA and the OFC/AMG nodes for target and distractor are updated with the same rules as the weight of the DD effect. This is again implemented to ensure that ME effects will decay after a certain amount of time (controlled by the tau constant) and will not persist indefinitely.

Finally, we need to account for the electrophysiological results from Kiss et al. (2007) through our model. As mentioned in Section 1, the main finding of Kiss et al.

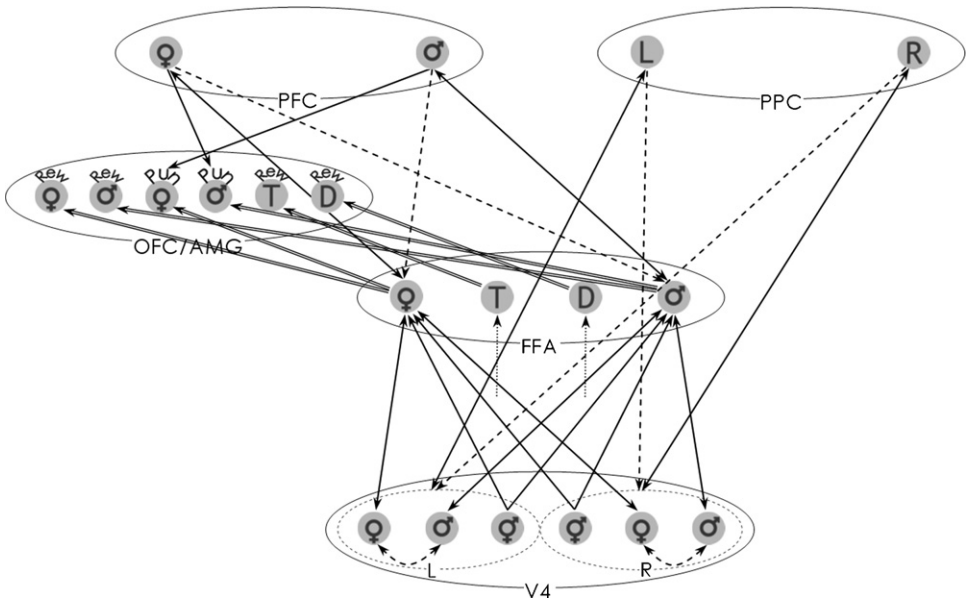


**Fig. 4.** The overall architecture of the DD model. Solid lines correspond to excitatory connections while dashed lines correspond to inhibitory connections. Double lines indicate connections with weight adaptation according to modified Hebbian described in main text. The blobs with biological gender-symbols correspond to face-gender-feature-sensitive neurons (V4) or face-gender-sensitive neurons (FFA, PFC, OFC). Blobs with letter L(R) correspond to left(right)-hemisphere-sensitive neurons while blobs with T(D) letters correspond to target(distractor)-coding neurons. OFC blobs with Rew(Pun) arched above them are coding for reward(punishment) of blob object. V4 module has left and right hemispheric components indicated by L and R dotted-line-defined groups. For details of nature of coding in various modules see main text.

(2007) was a correlation between the size of the N2pc and the level of subsequent DD in the bilateral face paradigm. We thus firstly need to generate a simulated N2pc by our model. As we stated earlier, the N2pc is the difference between the contralateral-to-target voltage measured over posterior scalp sites and the ipsilateral one, so to obtain the N2pc we must reproduce the two voltage waveforms and take the difference. The link between any ERP and the underlying neural activity is highly complex, and only few studies have attempted to reproduce ERPs from a neuronal model (David, Harrison, & Friston, 2005; David & Friston, 2003). It is, however, generally accepted that ERPs reflect mainly summated postsynaptic potentials (PSPs) (Nunez, Wingeier, & Silberstein, 2001). Our model uses simple graded-response neurons, so that the best approximation of the PSP is the membrane potential of the neuron(s). In a study of the N2pc that combined EEG and MEG measurements, Hopf et al. (2000) localised the N2pc in (and around) V4, while also showing that a magnetic equivalent of the N2pc contains an early component with parietal origins. Thus we take the sum of membrane potentials of V4 neurons for the 'contralateral to the target' component (i.e., where the target is actually processed) to be the contralateral voltage and the membrane potentials of the ipsilateral V4 neurons to be the ipsilateral voltage and, finally, take the difference to be the simulated N2pc. All things being equal, no difference is expected to exist between the left- and right-hemisphere activation in V4 when both sides have received an equal amount of stimulation from lower visual layers. What does make a difference is the positive feedback from PFC to the target-gender neuron and negative feedback to the distractor-gender neuron in FFA, which cascades back to the corresponding gender-feature-coding neurons in V4. Although this feedback is hemisphere-symmetric, the actual faces shown are asymmetric with respect to their gender and thus cause an asymmetric activation of the V4 neurons (that have already been target-gender biased). The neurons from the two hemispheric components of V4 project to the two hemispheric neurons of the PPC module which, in turn, feedback to the same-hemisphere V4 excitatorily and the opposite-hemisphere V4 inhibitorily causing a further differentiation of the V4 hemispheric activity. Thus, we propose, this PPC feedback signal could be the underlying source of the early parietal component of the magnetic equivalent of the N2pc as measured by Hopf et al. (2000). And it is this feedback signal that reinforces the imbalance of activity between the target-processing and the distractor-processing parts of the V4, manifesting as the N2pc difference wave.

We have thus related the PPC feedback (that occurs around 250 ms post-stimulus onset) with the creation of the N2pc in the V4 layer as measured by the occipito-parietal electrodes PO7 and PO8. This PPC feedback, as explained above, is a reflexive (exogenous) spatial attentional signal that is actually driven by the endogenous feature-based attentional feedback from the PFC to the FFA and, consequently, the V4 layer. We can thus change the effectiveness of the PPC attentional feedback (and so the size of the N2pc) by changing the parameters that control the PFC attentional feedback. More specifically, by increasing the weight of the PFC feedback inhibition to the distractor-gender FFA neuron we can obtain an increase of the N2pc size (as a result of larger V4 hemispheric imbalance and consequent PPC feedback). More PFC distractor inhibition, however, means more DD by means of the DD mechanism described above (through the PFC–OFC collaterals). This way we can link the size of the N2pc with the magnitude of DD as reported by Kiss et al. (2007). Furthermore, since PFC distractor inhibition is copied only onto the distractor punishment OFC node and not onto any target value nodes in the OFC, the distractor inhibition strength and, consequently, the magnitude of the N2pc, is disassociated with the value of the target and, thus, with the subsequent trustworthiness evaluations. This is in agreement with a further and crucial result from Kiss et al. (2007) that showed that higher or lower trustworthiness evaluations of the targets did not correlate with the preceding N2pc magnitude, as was the case for the distractors.

## 3. Results

The model described in the previous section deals explicitly only with the attentional selection phase of the two-face paradigm. We have modelled the modification of the values of the targets and distractors that occur *during* attentional selection under the assumption that these (modified) values would be the defining components of a formula used by evaluators in order to assess the trustworthiness of the face shown at the trustworthiness evaluation phase. In other words, the combined value of the reward and punishment nodes for the target and distractors, that may or may not have been altered during the attentional selection phase, would define the value of trustworthiness of the faces shown *after* attentional selection.

In order to best present the simulation results of our model of the DD effect we will focus on the reproduction of the experimental results from the study of Kiss et al. (2007). As mentioned above, Kiss and colleagues took electrophysiological measurements of the brain activity while participants performed the two-face bilateral task. These measurements are most useful for the validation of our model as they provide a link between the observed behaviour and the underlying neural activity, which is what our model is able to reproduce by means of the outputs of the neurons that comprise it. Through these results, however, we will also try to link the more complex and high-level behavioural results from both the Kiss et al. (2007) and the Goolsby et al. (2009) studies under the assumptions proposed and discussed in the previous section. Kiss et al. (2007) reported a link between the size of the N2pc extracted during the

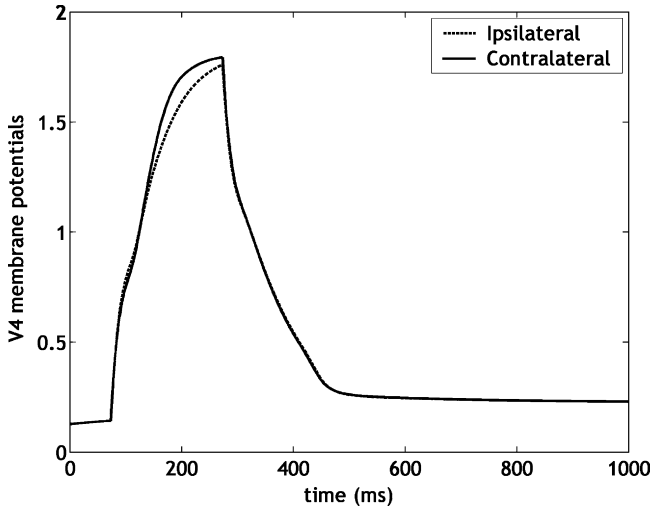


**Fig. 5.** Simulated N2pc for ventral attention feedback inhibition strength = 0.2.

attentional stage of the paradigm and the subsequent trustworthiness evaluations of the distractor faces. In our model, as described above, this link is achieved by means of PFC axon collaterals that transfer the PFC feedback inhibition to the value nodes in the OFC. Thus, varying the PFC feedback inhibition, on the one hand, leads to a variation of the observed N2pc and, on the other hand, a variation of the value of the distractors as coded in the OFC. This is illustrated in Figs. 5–7.

Figs. 5 and 6 show our model's N2pc equivalent, which is the summated membrane potentials of the contralateral and ipsilateral components of the V4 module (solid and dashed lines, respectively in the figures). These plots, however, do not correspond solely to the N2pc ERP (which should be a short peak around 250 ms) but show the timecourse of the V4 neurons' activations from the early bottom-up activations by lower-tier visual areas (starting at around 70 ms) to the later contributions from the PPC feedback. Note that the contralateral and the ipsilateral activations become differentiated from an early point (around 120 ms) due to the PFC ventral attention feedback with this difference being augmented a bit later (around 250 ms) by the PPC feedback. Thus, these plots present the fused effect of the early PFC feedback and the later PPC feedback which is why the resulting curves have a long duration. This fusion of high-level feedback activity is discussed in more detail in the following section. It must be noted, however, that the shape of the ipsilateral and contralateral activation curves plotted in Figs. 5 and 6 is defined to a large extent by the value of the model parameters. So, for instance, increasing the excitatory feedback from PFC would make the early difference (gap) between the ipsi- and contralateral curves even larger and, similarly, increasing the excitatory feedback from PPC would make the late difference of these curves larger. The latter can be achieved also by alternatively (or additionally) increasing the inhibitory feedback from the PPC. The strength of the lateral inhibition within V4 (male vs. female feature neurons) can also affect the size of the difference between the two curves as well as their overall shape. For the needs of our simulations, we have arbitrarily chosen a set of values for our parameters that yield reasonable results while allowing us to vary the crucial parameter of the PFC feedback inhibition strength in order to make the link with the OFC-based devaluation.

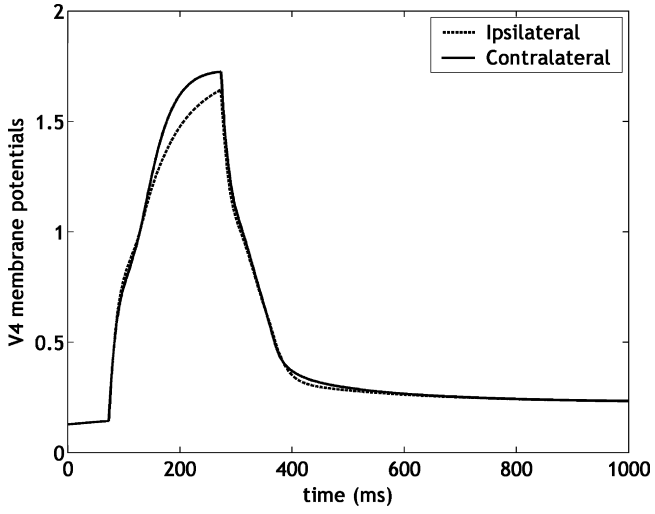


**Fig. 6.** Simulated N2pc for ventral attention feedback inhibition strength = 0.8.

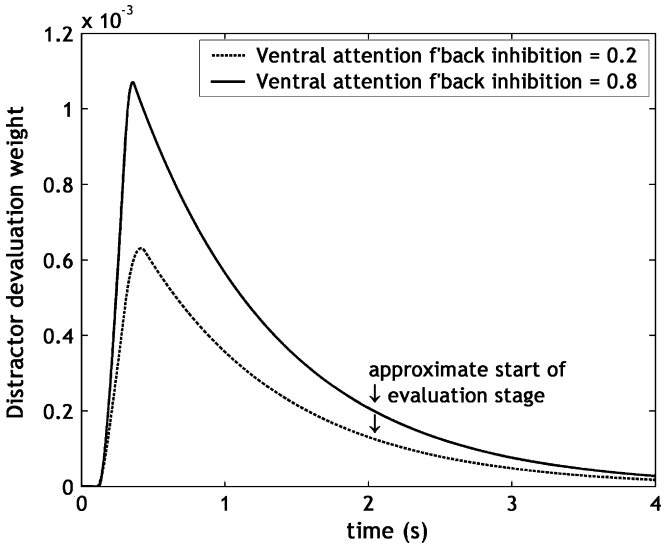


**Fig. 7.** Weight adaptation for distractor devaluation (DD).

Fig. 7 shows the change of the weight between the distractor FFA node (male or female depending on task instructions) and the distractor-punishment OFC node. As explained in the previous section, this weight change obeys a modified Hebbian learning rule that causes it to increase under the influence of the PFC feedback activation of the OFC node. The latter is communicated by the axon collateral of the PFC attention feedback onto the FFA distractor node. The weight change rate is governed by the time constant tau ($\tau$) so that for large values of tau the increase of the weight during stimulus input is slower (thus reaching at a smaller peak) but so is the weight decay. The latter (slow decay) would correspond to a longer lifetime of the DD effect. Experimental investigation of the lifetime of the DD effect is required to sharpen the time constant for the learning rule. With respect to the Kiss et al. (2007) results, we have plotted in Fig. 7 the change of the DD weight for two extreme values of the PFC feedback inhibition. The arrows indicate the value of the DD weight after attentional selection and a blank fixation period (that together take about 2 s), so, roughly, at the beginning of the trustworthiness evaluation stage. We hypothesise that in the condition where the distractor (or a novel face of the distractor gender) is presented again for trustworthiness evaluation, the value of the DD weight at that time will influence the outcome of the evaluation as explained in the beginning of this section. Thus, as shown in Fig. 7, when the PFC feedback inhibition is strong (0.8), the DD weight is larger than when the PFC feedback inhibition is weak (0.2), thus leading us to predict a lower distractor trustworthiness value. On the other hand, since the PFC distractor

**Table 1**
Simulation results for different values of ventral attention feedback inhibition strength.

| | Ventral attention feedback inhibition strength | | | |
|---|---|---|---|---|
| | 0.2 | 0.4 | 0.6 | 0.8 |
| Contralateral–ipsilateral activations | 0.035 | 0.045 | 0.060 | 0.085 |
| Distractor devaluation weight at evaluation | 1.309 | 1.338 | 1.555 | 2.075 |

inhibition has no effect on the OFC/AMG target nodes, the value of the target reward and punishment nodes will not depend on the strength of the PFC feedback inhibition and would be represented in Fig. 7 as a flat line (indicating no change under attention). It must be noted that although there is great variability between trials and between subjects in the two-face experiments both in the trustworthiness evaluations (shown if Fig. 2, averaged) and the N2pc waveforms, we have not explicitly tried to address this variability. Instead, the simulations run by our model were purely deterministic and aimed only to reproduce the observed correlation between grand averaged N2pc's and trustworthiness evaluations. We have assumed that the variation of the PFC feedback inhibition causes the observed variation of the N2pc; however, the source of this variation is not known. Trial-to-trial variations as well as subject-to-subject variations must play their role in the variation of the N2pc that, in the two-face paradigm, also correlates with the variation of the distractor evaluations. An example of the N2pc variability across subjects can be found in Luck (2005, p. 19). Table 1 lists the values of the simulated N2pc and the values of the DD weight for some intermediate values of the PFC feedback inhibition (0.4 and 0.6). These values indicate that the DD weight grows exponentially with the PFC feedback inhibition, which is expected due to the form of the DD weight change rule. These results should be taken as predictions from our model. Further experimental studies should offer validation or, alternatively, lead to the formulation of different adaptation rules and mechanisms. We note finally that we can consider the variations of lateral inhibitory strengths used in Table 1 as arising from variations in the subject's brains during a given trial, the variations being caused by modulations of the weight strengths from activities in other areas unrelated to the paradigm under investigation. Thus the data from Table 1 can be regarded as a particular implementation of subject variability behind the correlation between the N2pc level and the devaluation strength.

Fig. 8 shows the effect of ME on either the target or the distractor as represented by the increase of the weight from the FFA nodes to the corresponding positive value (reward) nodes in the OFC. The two plots correspond to either stimuli presented until response in the attentional stage (Goolsby et al., 2009: experiments 1 and 2) or stimuli presented for 200 ms (Goolsby et al., 2009: experiment 3). According to the weight adaptation formula described in the previous section, the weights increase as long as the stimuli are on (driven by the activations in the FFA) and decay when the stimuli are no longer presented. The value of the weight at the evaluation stage determines the amount of ME that the target or distractor receives as indicated in the figure by the arrows. The ME that both the target and the distractor receive during the attentional stage results in slightly boosted trustworthiness evaluations relevant to the evaluations of the novel faces. In the case of the distractors, particularly, the ME works antagonistically to the DD (as shown in Fig. 2 from Goolsby et al. (2009) and, so, the resulting trustworthiness evaluations given for the distractors are a balance between these two attention/emotion effects.

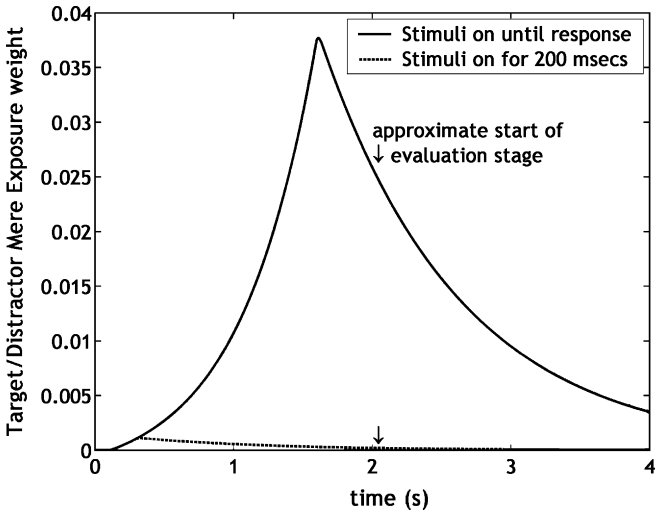


**Fig. 8.** Weight adaptation for mere exposure (ME).

## 4. Discussion

The aim of the present study was to investigate computationally possible neural mechanisms that give rise to the observed effects of attentional selection on emotional evaluations. To this end, we have developed a neural model that reproduces effectively the behavioural and neurophysiological results from recent studies of the so-called DD effect. We have adopted a limbic-based interpretation of the DD effect that has attentional inhibition being copied onto value-coding nodes in the OFC and amygdala, causing a reduction of the distractor value in these brain regions. This distractor value modulation then causes the differential target/distractor trustworthiness evaluations. Such evaluations are known to activate several limbic structures, such as the OFC and the amygdala (Winston, Strange, O'Doherty, & Dolan, 2002), so it is reasonable to assume that the current value of a stimulus is being read-off the OFC and the amygdala in order to generate the trustworthiness evaluation. However, whether the modulation of distractor value occurs in parallel with the attentional selection (by means of PFC axon collaterals, as proposed here) is unclear. It is conceivable that DD operates in a way similar to the perceptual fluency hypothesis of ME, only in inverse. That is, attentional inhibition causes perceptual anti-fluency onto the distractor which is then misattributed, at the evaluation stage, as reduced trustworthiness. This interpretation would suggest that attentional selection, though it has affective consequences, does not directly interact with the emotional system and, thus, is a purely cognitive phenomenon. One way to help delineate the underlying mechanism for the DD effect is by means of brain imaging. If the DD effect relies on activation/modulation of the OFC and the amygdala during the attentional selection stage (as hypothesised by our model) differential modulatory effects of these structures by attentional selection should be observed in a suitably designed fMRI study. If, on the other hand, fMRI does not reveal any modulation of OFC/amygdala during the attentional selection stage then a more cognitive explanation would be more apposite.

In our model, ME is explained by an automatic elevation of the value of stimuli presented during the attentional selection stage that is encoded in the OFC/AMG. This results in those stimuli being rated as higher in the trustworthiness evaluation stage relative to

novel stimuli that have not had this prior value elevation. A natural question that arises from this proposed mechanism is why should simple presentation (without any related reward) elevate the value of the presented stimuli. Currently, no studies have been carried out to further explore this interpretation of ME; it is however possible to obtain experimental confirmation for the conditioning model of ME using brain imaging (such as fMRI) of the OFC and the amygdala, as well as with the well-established single-cell recordings methodology used for animal conditioning studies (Schoenbaum, Chiba, & Gallagher, 1999). If confirmation for the conditioning model is not obtained, then more cognitive accounts of the ME (such as the perceptual fluency hypothesis) must be considered alternatively.

Our model of attentional influences on emotional evaluations crucially depends on the assumption that our learning rules underpin the DD and the ME effects. In general, synaptic plasticity comes in many varieties and occurs over a large range of timescales. The most well-studied type of plasticity is that which results from long-term depression and potentiation (LTD and LTP) of synapses and is thought to underlie the formation of associations in learning and memory (Bi & Poo, 2001). However, this type of plasticity lasts for hours (and often much longer) and is thus not relevant to our more subtle and short-lived effects (DD and ME). We are interested in shorter term plasticity, here, such as that occurring over a few seconds or minutes. Individual synapses of single neurons can be facilitating or depressing depending on a number of variables such as the pre-synaptic activity. Given that single neurons usually form a large number of synapses that themselves can behave diversely and considering that plasticity normally results from the interaction of large populations of neurons, the possibilities of learning rules that could govern the formation of plastic connections in such neuronal populations are numerous (see Abbott and Regehr (2004) for a review). Common forms of short-term plasticity include facilitation, augmentation, and post-tetanic potentiation and are usually the result of residual elevation in pre-synaptic $[Ca^{2+}]$ (Zucker & Regehr, 2002). We have proposed a form of short-term plasticity that is a variant of the classical Hebbian rule with a decay term that lets the connection become weak again when no longer reinforced. The existence of such forms of synaptic plasticity has not yet been experimentally established, but it has been shown to yield robust results when employed in a neural model of working memory (Sandberg, Tegner, & Lansner, 2003). Synaptic decay according to this rule is controlled by the time constant tau, which can then be fitted according to what experimental investigation would reveal the lifetime of DD and ME to be. More generally, further experimental studies, carefully designed, are needed to delineate the exact nature of the learning laws that underpin the DD and ME effects.

Regarding the ME lifetime, in particular, the plots of the weight timecourse for the two stimulus duration cases in Fig. 8 suggest a very large difference of ME depending on stimulus duration. The behavioural results from Goolsby et al. (2009) (reproduced in Fig. 2) do show a difference in the amount of ME that the target and detractor receive in the two different stimulus duration conditions (experiments 2 and 3) but it appears to be rather small (and certainly not statistically significant). Thus our resulting difference may be unrealistic and not observable experimentally. A plausible explanation for this discrepancy is that there is saturation of the reward weight due to stimulus exposure after a certain exposure time. Thus any stimulus duration beyond the weight saturation point will not cause any further increase on the weight and, consequently, any more ME.

The mechanism for the DD used in the model is one that is based on weight adaptation. Alternatively, an activation mechanism for the DD that is based on the continuous activation of the distractor OFC punishment node by the PFC target goal node can be used. This is in fact already included in our adaptation mechanism, only in the latter it is used to modulate the adaptation of the weight between the FFA and the OFC distractor nodes. An obvious limitation of an activation-based DD mechanism is that even with strong recurrence in the PFC to sustain this activation, it cannot really go on beyond a couple of seconds from offset of the attention PFC goal node. This poses great constraints on the lifetime of the DD, and experimental investigation is needed to confirm the plausibility of such a mechanism. Another limitation posed by the use of an activation mechanism for the DD is the need for dedicated representations of the male and female faces in the OFC/AMG. The existence of such representations can only be conjectured as no experimental evidence supports it to date. The adaptation-based mechanism, on the other hand, encodes the DD in the value of the weight between the face gender representations in the FFA and the respective nodes in the OFC/AMG. However, although in the present model we have indeed used dedicated representations for the face genders in the OFC/AMG, the adaptation-based mechanism allows us to, alternatively, use generic reward and punishment nodes in the OFC/AMG (that are better supported experimentally) with the strength of the connections to specific object nodes in the FFA defining the value of those objects.

We propose a mechanism for the generation of the N2pc that can be tested experimentally in a variety of conditions as prompted by variations of our model's parameters. In our model, the N2pc is merely a corollary of the imbalance of activity between the two hemispheric components of V4 which is mainly caused by the feature-based PFC attentional feedback and reinforced by the spatial PPC attentional feedback. This latter feedback would correspond to the early component of the magnetic equivalent of the N2pc (Hopf et al., 2000) and it involves both ipsilateral excitation and contralateral inhibition. Therefore the N2pc, measured here as the difference of the ipsilateral and contralateral sums of membrane potentials of the neurons in V4, reflects both excitatory and inhibitory effects.

In order to model the link between the size of the N2pc and the level of devaluation of distractors as reported in Kiss et al. (2007) we have assumed that attentional feedback from the PFC to the FFA also has an excitatory and an inhibitory component (like the PPC). When the PFC node that corresponds to the target gender becomes active under the task instructions, it feeds back excitatorily to the target-gender nodes in FFA and inhibitorily to the distractor-gender nodes. Via the feedback connections between the FFA and V4, this PFC feedback causes V4 target-gender-feature nodes to be favourably biased against distractor-gender-feature nodes which, as described above, leads to the activity imbalance that manifests eventually as the N2pc. The division of the PFC signal into an excitatory and inhibitory component allows us to manipulate the respective strengths of these components separately. Since Kiss et al. (2007) observed only the distractors' trustworthiness evaluations to covary with the N2pc magnitude (and not the targets'), we linked the inhibitory component of the PFC feedback (the one directed to the distractor FFA node) with the signal that increases the distractor punishment in the OFC. This was achieved by means of axon collaterals as explained above. Consequently, increase of the strength of the PFC distractor causes an increase of distractor punishment. By increasing the inhibitory feedback from the PFC to the FFA, however, the activity imbalance in the V4 layer is increased leading to a stronger PPC feedback in favour of the hemisphere the target is being processed in and, finally, as a larger N2pc.

We close the discussion by proposing a set of predictions generated by the model that refer to recent and possible future experimental studies. Firstly, fMRI studies of attention have reported task-related preparatory BOLD activity in the areas that would subsequently be activated by the task-relevant stimuli (Chawla, Rees, & Friston, 1999; Cristescu, Devlin, & Nobre, 2006;

Giesbrecht et al., 2003; Kastner, Pinsk, De, Desimone, & Ungerleider, 1999). This preparatory activity presumably reflects some top-down pre-activation/sensitisation of the areas that would process the relevant stimuli upon presentation, in order to make the overall system more efficient (a correlation between the magnitude of the preparatory and subsequent task performance has in fact been reported in Stern and Mangels (2006) – see also Sylvester, d'Avossa, and Corbetta (2006). Our model assumes such a preparatory signal is dispatched from the PFC in order to enhance the target category while inhibiting the distractor category in the FFA. Moreover, the inhibitory signal is 'copied' by means of axon collaterals onto the OFC punishment nodes causing distractor devaluation. Thus, our model would predict an early (pre-presentation) activation of the FFA (reflecting endogenous attentional feedback) as well as an equal (or covarying) BOLD activity in the OFC (reflecting the DD signal as a 'copy' of the attentional feedback). The BOLD activity in the OFC would also be expected according to our model to covary with the subsequent distractors' trustworthiness ratings. The two-face task would of course have to be adapted to the requirements of an fMRI study using stimuli whose activity could be differentiated with fMRI (e.g. faces vs. houses) and intervals that are sufficiently long and jittered to enable individuation of hemodynamic responses to the separate trial events (e.g., see Lepsien and Nobre, 2007). It should be noted that the mechanism of PFC pre-activation feedback and its collateral copy on the OFC that was implemented in the model in order to account for the behavioural results of the DD studies is not itself a model prediction *per se*. This mechanism, however, has apparent implications regarding the underlying neural activity that can be rigorously tested through suitable imaging experiments and as such represent important experimental predictions.

Another prediction from our model is a break-up of the link between the amplitude of the N2pc and the level of DD that would arise from disabling the PPC module. Disabling the PPC module in the model would diminish the simulated N2pc while distractor devaluation would still be brought about via the 'copy' of the PFC inhibitory feedback onto the OFC/AMG module. Experimentally, the PPC can be disrupted using TMS as explored in Fuggetta, Pavone, Walsh, Kiss, and Eimer (2006). The authors showed that when TMS was applied to the right PPC not only were response times to targets during conjunction search delayed, but the early part of the N2pc was eliminated. TMS application to a control site on the other hand did not have a behavioural effect nor cause a change the N2pc. Our model would thus predict that if TMS was applied to the PPC in the attentional stage of the two-face paradigm, there would still be a strong DD effect without, however, a covarying N2pc as reported in Kiss et al. (2007). This is illustrated in Fig. 9 where the effect of TMS

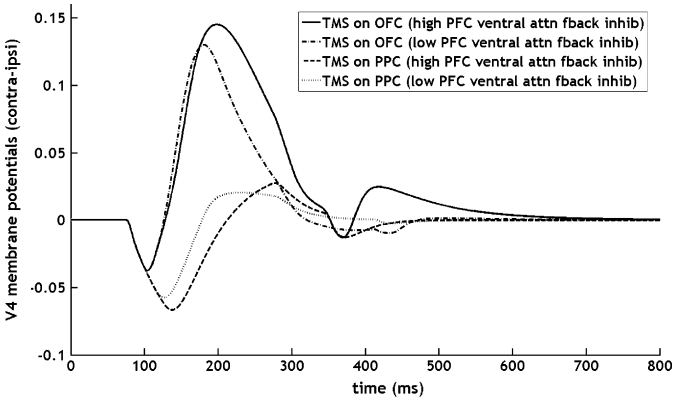


**Fig. 9.** Model simulation of TMS application on the PPC compared with TMS applied to the OFC. TMS on PPC almost extinguishes the N2pc whereas TMS on the OFC does not have an effect on the N2pc.

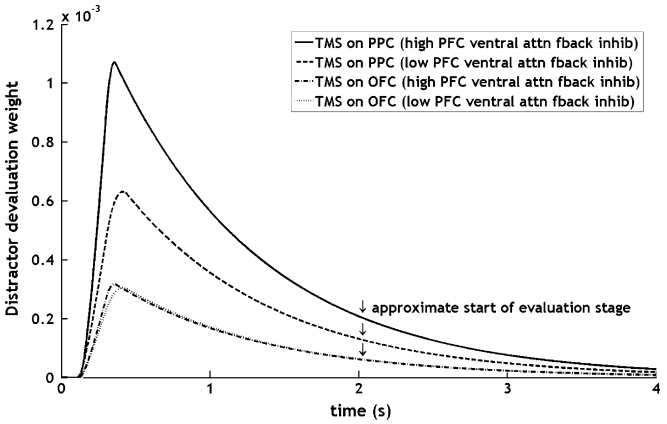


**Fig. 10.** Model simulation of TMS disruption of the OFC compared with TMS disruption of the PPC. TMS disruption of the OFC eliminates the difference between high and low PFC ventral attentional inhibition (and thus no DD observed) whereas disruption of the PPC does not have an effect on the DD magnitude.

applied to the PPC is simulated by disconnecting the PPC module from the network.

Conversely, our model would predict that disruption of the OFC, such as by TMS application to the OFC (although this might not be possible because the location of the OFC is not very accessible for TMS stimulation), would result in loss of any DD and a complete de-coupling of the N2pc magnitudes with the following trustworthiness ratings for distractors. That is to say that although the top-down attentional signal would help generate an N2pc in the fashion proposed by our model, the collateral signal descending to the OFC would be rendered ineffective due to the OFC disruption. TMS disruption of the OFC was simulated by reducing the weight of the Hebbian component in the DD-related weight update equation from 1 to 0.5:

$$\tau \frac{dw}{dt} = -w + 0.5 * (\text{input})(\text{output})$$

and is illustrated in Fig. 10.

Another prediction from our model is related to the expected outcome of varying the target–distractor similarity. In the Goolsby et al. (2009) paradigm varying the target–distractor similarity would correspond to making the male and female faces more or less distinguishable with respect to their gender, thus making the gender classification task easier or harder. In our model, such variation can be simulated by varying the ratio of male/female to androgynous nodes' weights in the V4 module. Experimentally, a new face set would have to be constructed that would contain faces of both distinct gender and ambiguous ones. Thus, the target–distractor similarity would be directly manipulated experimentally. Following Conci, Gramann, Muller, and Elliott (2006) we would expect the amplitude of the N2pc (as reported in Kiss et al., 2007) to vary inversely to the similarity of the target and distractor (the difficulty of the attentional task). Thus, the more androgynous the faces, the smaller the N2pc obtained in the electrophysiological experiment. In our model, increasing the androgynous nodes' activation would generate a reduced (simulated) N2pc as the distractors become more active in V4 and FFA owing to their common features with the targets. This is consistent with Conci et al. (2006) that show that the reduction of the N2pc to the targets when the distractor share more features with the target is due to an increase of the ipsilateral to the target voltage (which corresponds to the hemisphere that the distractor is processed). The model's simulation of the N2pc reduction caused by more androgynous faces is illustrated in Fig. 11.

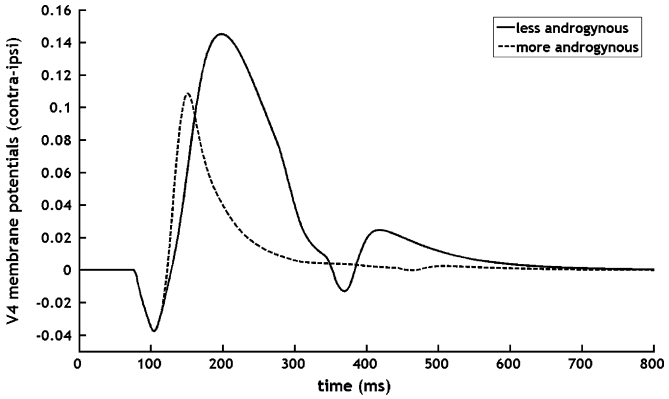


**Fig. 11.** The reduction of the N2pc as brought about by increasing gender similarity of the target and distractor faces, consistent with Conci et al. (2006).

However, as it stands, our model does not relate the difficulty of the task with the degree of devaluation that distractors suffer as the latter is decided at the beginning of the trial and stays fixed thereafter. The experimental prediction from our model would thus be that, although more androgynous distractors would yield a smaller N2pc (as reflecting task difficulty), the devaluation would not depend on that. Although this experimental prediction follows straightforwardly from our current model, it may be unrealistic and as such it would suggest our model is limited and should be extended by the addition of a monitor module that would monitor the attainment of the task goal (i.e., find the male/female face) and would thus be sensitive to the increased difficulty associated with the increased similarity between the two faces. This monitor would then activate the prefrontal attention goal sites further so that more resources be allocated to the task. This would, in turn, cause further collateral distractor devaluation in the OFC/AMG sites by means of the proposed 'copy' of inhibition. The experimental prediction of this scenario would thus be an increase of the distractor devaluation with an increase in target–distractor similarity (and the associated suggested decrease in the N2pc amplitude).

In order to investigate the influence of task difficulty on the strength of the attentional preparatory signal discussed above, a variant of the two-face paradigm with an additional cue before the attention task indicating whether the trial would be hard (more androgynous faces) or easy (more gender-distinct faces) could be explored. The cue could belong to any modality (such as visual, auditory etc.) and would cause participants to prepare for the ensuing task accordingly. Our model would thus predict that if the cue indicated a hard trial then the attentional preparatory signal would be stronger registering as a stronger BOLD signal in the PFC as well as a collateral strong BOLD signal in the OFC. The latter would have the behavioural consequence of strong distractor devaluation (owing to the increased inhibition applied in order to perform the attentional task). If, on the other hand, the cue indicated an easy trial the BOLD signal in the PFC and OFC would be weaker (i.e. reduced relative to the hard trials) followed by a weaker DD.

## 5. Conclusion

The present study has provided a computational investigation of recently reported influences of attentional selection on emotional evaluations. As such, this study complements previous modelling studies of the reciprocal link, that is, of emotional effects on attentional selection. This study has focused on a phenomenon termed 'distractor devaluation' after findings showing distractors being consistently rated less positively in emotional evaluations, following attentional selection in visual search paradigms, relative to targets. A neural model was developed in order to account for these findings and provide a plausible mechanism that underlies the observed effects. The model comprises a set of modules, each assigned a specific function (such as feature-coding, feature or spatial attention, value coding and so on) and a proposed siting in the brain. Furthermore, the modules of the model are connected to one another according to known neuroanatomical and functional links, as well as conjectured links that help explain the DD results but merit further experimental support and confirmation. The model successfully reproduces the various behavioural features of DD as well as reported systematic covariations of an electrophysiological measure of attention and the magnitude of DD.

The model also provides a set of testable hypotheses whose validation would provide both a stronger justification of the model details as well as help give a better understanding of the manner in which attention can affect emotional valuation of stimuli. More specifically, we have hypothesised that:

(i) PFC axon collaterals transfer attentional inhibition onto value-encoding nodes in the OFC and the amygdala thus giving rise to the observed distractor devaluation.
(ii) Mere exposure is caused by the automatic increase of value occurring during initial presentation of the faces.
(iii) The modification of target/distractor value during attentional selection occurs by means of weight adaptation and, specifically, via a modified Hebbian rule that decays with time.
(iv) The N2pc arises when an initial hemispheric imbalance of activity in V4 gets reinforced by PPC feedback.
(v) The PFC attentional feedback has an excitatory and an inhibitory component that may be dissociable and, thus, separately manipulated.

The model we have developed is based on a more general model of attention, the CODAM model (Fragopanagos et al., 2005; Taylor, 2003) where CODAM stands for Corollary Discharge of Attention Movement. CODAM is an extension of the interesting recent applications of such ideas to motor control now to the field of attention control. CODAM is thus being considered in the DD paradigm we have discussed in the case of the interaction of bottom-up and top-down attention. Both forms of attention were originally included the CODAM model, although most simulations have been based on the endogenous form (Korsten, Fragopanagos, Hartley, Taylor, & Taylor, 2006). CODAM includes the majority of other models of attention control, especially the influential 'biased competition' model of Desimone and Duncan (1995) mentioned at the beginning. Thus the exercise we have performed in the simulation can be regarded as an exercise in testing CODAM further in the case of the presence of both exogenous and endogenous attention in interaction.

## Acknowledgements

This research is supported by Integrative Analysis of Brain and Behaviour (IABB) initiative grant BBS/B/16178 from the Biotechnology and Biological Sciences Research Council (BBSRC), UK. Principal investigators are Martin Eimer, Anna Christina Nobre, Jane E. Raymond, Kimron Shapiro, and John G. Taylor.

## Appendix A. Appendix

In this appendix we give the mathematical equations which describe the firing activity and synaptic dynamics of the model as well as a table of the selected values for the parameters of the model.

### A.1. The neuron equations

Each neuron in our simulation is a simple graded-response neuron with a membrane equation:

$$\tau \frac{dV}{dt} = -V + \rho I,$$

where $V$ is the neuron membrane potential, $\tau$ is the neuron time constant, $I$ is the injected current from the neuron's various connections and $\rho$ is a constant that regulates the voltage to current ratio.

The output of each neuron is the positive part of a saturating sigmoidal non-linearity:

$$\text{Output} = f(V(t)),\ \ f(x) = \left[\frac{1}{1+e^{-\frac{x-\vartheta}{T}}} - \frac{1}{2}\right]_+,$$

where $\theta$ is the threshold and $T$ is the noise temperature.

### A.2. The modules' equations

#### A.2.1. Input

At each trial of the two-face paradigm, a face is presented for 200 ms on either side of the fixation cross and each face can be either male or female. This display is represented in our model's Input module (not shown in Fig. 3) by two unit amplitude pulses of duration $t_s$ = 200 ms. Two genders and two hemifields yield four possible configurations that are indexed by the $i$ and $k$ indices in the following equation:

$$I_{ik}^{IN} = \begin{cases} 1 & \text{for } 0 < t < t_s \\ 0 & \text{for } t > t_s \end{cases},$$

where $i \in \{M,F\}$ and $k \in \{L,R\}$, with $\begin{matrix} M: Male \\ F: Female \end{matrix}$ and $\begin{matrix} L: Left \\ R: Right \end{matrix}$.

#### A.2.2. V4

The V4 module contains six neurons: two gender-feature-specific and one non-specific, for each hemisphere. Each gender-feature-specific neuron in the module is directly activated by the same-gender face stimulus from the Input module while the non-gender-specific (hermaphrodites) are activated by both gender face inputs. The gender-feature-specific neurons also compete with each other within each hemisphere by means of lateral inhibition. Finally, the gender-feature-specific neurons receive reinforcing feedback from the same-gender neurons in the FFA module. The equation for the gender-feature-specific neurons is thus:

$$\tau_{V4} \frac{dV_{ik}^{V4}(t)}{dt} = -V_{ik}^{V4}(t) + \rho\{I_{il}^{IN}(t - t_{V4-IN}) + I_i^{V4-FFA}(t) + I_k^{V4-PPC_{EXC}}(t) - I_l^{V4-PPC_{INH}}(t) + I_{ik}^{V4_{LAT}}(t)\},$$

$$I_i^{V4-FFA}(t) = w_i^{V4-FFA} f(V_i^{FFA}(t)),$$
$$I_k^{V4-PPC_{EXC}}(t) = w_k^{V4-PPC_{EXC}} f(V_k^{PPC}(t)),$$
$$I_l^{V4-PPC_{INH}}(t) = w_l^{V4-PPC_{INH}} f(V_l^{PPC}(t)),$$
$$I_{ik}^{V4_{LAT}}(t) = w_{jk}^{V4_{LAT}} f(V_{jk}^{V4}(t)),$$

where $i, j \in \{M,F\}$, $i \neq j$ and $k, l \in \{L,R\}$, $k \neq l$, with $\begin{matrix} M: Male \\ F: Female \end{matrix}$ and $\begin{matrix} L: Left \\ R: Right \end{matrix}$.

The equation for the hermaphrodite V4 neurons is similar to that for the male/female neurons without the FFA feedback and the lateral inhibition terms:

$$\tau_{V4} \frac{dV_{Hk}^{V4}(t)}{dt} = -V_{Hk}^{V4}(t) + \rho\left\{\sum_i I_{il}^{IN}(t - t_{V4-IN}) + I_k^{V4-PPC_{EXC}}(t) - I_l^{V4-PPC_{INH}}(t)\right\},$$

with $H$ : *Hermaphrodite*.

#### A.2.3. FFA

The FFA module contains two gender-specific neurons and two neurons that respond to the specific identity of the target and the distractor. The gender-specific neurons are driven by the same-gender feature-specific neurons (from both hemispheres) of the V4 module described above as well as by both hermaphrodite V4 neurons. Furthermore, the gender-specific neurons receive feedback attentional signals from the PFC module, excitatory from the same-gender neurons and inhibitory from the opposite-gender neurons. Thus the equations for the gender-specific neurons become:

$$\tau_{FFA} \frac{dV_i^{FFA}(t)}{dt} = -V_i^{FFA}(t) + \rho\{I_i^{FFA-V4_{GEN}}(t - t_{FFA-V4}) + I_i^{FFA-V4_{NONGEN}}(t - t_{FFA-V4}) + I_i^{FFA-PFC_{EXC}}(t) + I_j^{FFA-PFC_{INH}}(t)\}$$

$$I_i^{FFA-V4_{GEN}}(t) = \sum_k w_{ik}^{FFA-V4_{GEN}} f(V_{ik}^{V4}(t)),$$

$$I_i^{FFA-V4_{NONGEN}}(t) = \sum_k w_{Hk}^{FFA-V4_{NONGEN}} f(V_{Hk}^{V4}(t)),$$

$$I_i^{FFA-PFC_{EXC}}(t) = w_i^{FFA-PFC_{EXC}} f(V_i^{PFC}(t)),$$
$$I_j^{FFA-PFC_{INH}}(t) = w_j^{FFA-PFC_{INH}} f(V_j^{PFC}(t)),$$

where $i, j \in \{M,F\}$, $i \neq j$ and $k \in \{L,R\}$, with $\begin{matrix} M: Male \\ F: Female \end{matrix}$ and $\begin{matrix} L: Left \\ R: Right \end{matrix}$.

Finally, the target/distractor identity FFA neurons are trivially activated directly from the input thus being described by the following equation:

$$\tau_{FFA} \frac{dV_m^{FFA}(t)}{dt} = -V_m^{FFA}(t) + \rho I_m^{FFA-IN}(t - t_{FFA-IN}),$$

$$I_m^{FFA-IN} = \begin{cases} 1 & \text{for } 0 < t < t_s \\ 0 & \text{for } t > t_s \end{cases},$$

where $m \in \{T,D\}$, with $\begin{matrix} T: Target \\ D: Distractor \end{matrix}$.

#### A.2.4. PPC

The PPC module contains two neurons, one responding to left-hemisphere stimuli and one to right-hemisphere stimuli. These neurons are driven by the corresponding hemisphere V4 neurons (three on each hemisphere). Thus the equation for the PPC neurons is:

$$\tau_{PPC} \frac{dV_k^{PPC}(t)}{dt} = -V_k^{PPC}(t) + \rho I_k^{PPC-V4}(t),$$

$$I_k^{PPC-V4}(t) = \sum_i w_{ik}^{PPC-V4} f(V_{ik}^{V4}(t)),$$

where $k \in \{L,R\}$ and $i \in \{M,F,H\}$, with $\begin{matrix} L: Left \\ R: Right \end{matrix}$ and $\begin{matrix} M: Male \\ F: Female \end{matrix}$ and $H$: *Hermaphrodite*

*A.2.5. PFC*

Two neurons comprise the PFC module representing the goal of the task, i.e. 'find the male/female face'. These neurons are gender specific (like the FFA neurons) and one of them – depending on the task instructions – becomes active and stays so until the attentional task is completed. The equations for the PFC neurons are thus:

$$\tau_{PFC}\frac{dV_i^{PFC}(t)}{dt} = -V_i^{PFC}(t) + \rho I^{ON}(t),$$

$$I^{ON} = \begin{cases} 1 & \text{for } 0 < t < t_{ATT} \\ 0 & \text{for } t > t_{ATT} \end{cases}, \quad t_{ATT} \simeq 1\,\text{s},$$

where $i \in \{M,F\}$, with $\begin{matrix} M : Male \\ F : Female \end{matrix}$.

*A.2.6. OFC*

The OFC module contains six neurons, four for face-gender value coding and two for the target and distractor value coding. Each face-gender has a reward and a punishment neuron coding for positive and negative value respectively, while the target and distractor only have a reward neuron in order to encode the mere exposure effect (see main text). All OFC neurons are connected to their FFA counterparts via the modified Hebbian weight described in the main text. Furthermore, the punishment neurons for the two genders are being fed by the PFC opposite-gender neurons in order to achieve distractor devaluation via the PFC axon collaterals of the attentional distractor inhibition (see main text for details of this mechanism). So the equations for the punishment face-gender neurons are:

$$\tau_{OFC}\frac{dV_i^{OFC_{GENPUN}}(t)}{dt} = -V_i^{OFC_{GENPUN}}(t) + \rho\left\{I_i^{OFC_{GENPUN}-FFA}(t) + I_j^{OFC_{GENPUN}-PFC}(t)\right\},$$

$$I_i^{OFC_{GENPUN}-FFA}(t) = w_i^{OFC_{GENPUN}-FFA} f(V_i^{FFA}(t)),$$

$$I_j^{OFC_{GENPUN}-PFC}(t) = w_j^{OFC_{GENPUN}-PFC} f(V_j^{PFC}(t)),$$

$$\tau_{DD}\frac{dw_i^{OFC_{GENPUN}-FFA}(t)}{dt} = -w_i^{OFC_{GENPUN}-FFA}(t) + \left(I_i^{OFC_{GENPUN}-FFA}(t) + I_j^{OFC_{GENPUN}-PFC}(t)\right) f \times \left(V_i^{OFC_{GENPUN}}(t)\right).$$

The equations for the reward face-gender neurons in the OFC are:

$$\tau_{OFC}\frac{dV_i^{OFC_{GENPUN}}(t)}{dt} = -V_i^{OFC_{GENPUN}}(t) + \rho I_i^{OFC_{GENPUN}-FFA}(t),$$

$$I_i^{OFC_{GENPUN}-FFA}(t) = w_i^{OFC_{GENPUN}-FFA} f(V_i^{FFA}(t)),$$

$$\tau_{REW}\frac{dw_i^{OFC_{GENPUN}-FFA}(t)}{dt} = -w_i^{OFC_{GENPUN}-FFA}(t) + I_i^{OFC_{GENPUN}-FFA}(t) f\left(V_i^{OFC_{GENPUN}}(t)\right),$$

where $i, j \in \{M,F\}$, $i \neq j$ with $\begin{matrix} M : Male \\ F : Female \end{matrix}$.

The equations for the target/distractor reward neurons are:

$$\tau_{OFC}\frac{dV_m^{OFC_{T/DREW}}(t)}{dt} = -V_m^{OFC_{T/DREW}}(t) + \rho I_m^{OFC_{T/DREW}-FFA}(t)$$

$$I_m^{OFC_{T/DREW}-FFA}(t) = w_m^{OFC_{T/DREW}-FFA} f(V_m^{FFA}(t)),$$

$$\tau_{ME}\frac{dw_m^{OFC_{T/DREW}-FFA}(t)}{dt} = -w_m^{OFC_{T/DREW}-FFA}(t) + I_m^{OFC_{T/DREW}-FFA}(t) f\left(V_m^{OFC_{T/DREW}}(t)\right),$$

where $m \in \{T,D\}$, with $\begin{matrix} T : Target \\ D : Distractor \end{matrix}$.

Finally, we present the values of the parameters in the equations above in the following table. The motivation for the choice of these values is discussed in the main body of the paper.

*A.3. Table of parameters*

| | |
|---|---|
| $\rho$ | 4.5 |
| $\theta$ | 0.01 |
| $T$ | 1 |
| $\tau_{V4}$, $\tau_{FFA}$, $\tau_{PPC}$, $\tau_{PFC}$, $\tau_{OFC}$ | 20 ms |
| $\tau_{DD}$, $\tau_{REW}$, $\tau_{ME}$ | 1 s |
| $\tau_{V4-IN}$ | 70 ms |
| $t_{FFA-V4}$ | 20 ms |
| $t_{FFA-IN}$ | 90 ms |
| $w_i^{V4-FFA}$ | 0.4 |
| $w_k^{V4-PPC_{EXC}}$ | 0.2 |
| $w_k^{V4-PPC_{INH}}$ | 0.2 |
| $w_{jk}^{V4_{LAT}}$ | 0.4 |
| $w_{ik}^{FFA-V4_{GEN}}$ | 1 |
| $w_{ik}^{FFA-V4_{NONGEN}}$ | 1 |
| $w_i^{FFA-PFC_{EXC}}$ | 0.2 |
| $w_{ik}^{PPC-V4}$ | 1 |
| $w_j^{FFA-PFC_{INH}}$, $w_j^{OFC_{GENPUN}-PFC}$ | varied (see text) |

## References


Abbott, L. F., & Regehr, W. G. (2004). Synaptic computation. *Nature*, *431*, 796–803.

Arrington, C. M., Carr, T. H., Mayer, A. R., & Rao, S. M. (2000). Neural mechanisms of visual attention: Object-based selection of a region in space. *Journal of Cognitive Neuroscience*, *12*(Suppl. 2), 106–117.

Bi, G., & Poo, M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience*, *24*, 139–166.

Bornstein, R. F., & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition*, *12*, 103–128.

Cavada, C., Company, T., Tejedor, J., Cruz-Rizzolo, R. J., & Reinoso-Suarez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cerebral Cortex*, *10*, 220–242.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*, 671–676.

Conci, M., Gramann, K., Muller, H. J., & Elliott, M. A. (2006). Electrophysiological correlates of similarity-based interference during detection of visual forms. *Journal of Cognitive Neuroscience*, *18*, 880–888.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews of Neuroscience*, *3*, 201–215.

Cristescu, T. C., Devlin, J. T., & Nobre, A. C. (2006). Orienting attention to semantic categories. *Neuroimage*, *33*, 1178–1187.

David, O., & Friston, K. J. (2003). A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *Neuroimage*, *20*, 1743–1755.

David, O., Harrison, L., & Friston, K. J. (2005). Modelling event-related responses in the brain. *Neuroimage*, *25*, 756–770.

Deco, G., & Rolls, E. T. (2003). Attention and working memory: A dynamical model of neuronal activity in the prefrontal cortex. *European Journal of Neuroscience*, *18*, 2374–2390.

Deco, G., & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. *Visual Cognition*, *8*, 118–139.

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

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & Psychophysics*, *63*, 1004–1013.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*, 225–234.

Fenske, M. J., & Eastwood, J. D. (2003). Modulation of focused attention by faces expressing emotion: Evidence from flanker tasks. *Emotion*, *3*, 327–343.

Fenske, M. J., Raymond, J. E., Kessler, K., Westoby, N., & Tipper, S. P. (2005). Attentional inhibition has social-emotional consequences for unfamiliar faces. *Psychological Science*, *16*, 753–758.

Fenske, M. J., Raymond, J. E., & Kunar, M. A. (2004). The affective consequences of visual attention in preview search. *Psychonomic Bulletin Review*, *11*, 1055–1061.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, *130*, 681–700.

Fragopanagos, N., Kockelkoren, S., & Taylor, J. G. (2005). A neurodynamic model of the attentional blink. *Brain Research Cognitive Brain Research*, *24*, 568–586.

Freese, J. L., & Amaral, D. G. (2005). The organization of projections from the amygdala to visual cortical areas TE and V1 in the macaque monkey. *Journal of Comparative Neurology*, *486*, 295–317.

Fuggetta, G., Pavone, E. F., Walsh, V., Kiss, M., & Eimer, M. (2006). Cortico-cortical interactions in spatial attention: A combined ERP/TMS study. *Journal of Neurophysiology*, *95*, 3277–3280.

Giesbrecht, B., Woldorff, M. G., Song, A. W., & Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage*, *19*, 496–512.

Goolsby, B., Raymond, J. E., Silvert, L., Kiss, M., Fragopanagos, N., Taylor, J. G., Eimer, M., Nobre, A. C., & Raymond, J. E. (2009). Feature-based inhibition underlies the affective consequences of attention. *Visual Cognition*, *17*(4), 500–530.

Hamker, F. H. (2004). A dynamic model of how feature cues guide spatial attention. *Vision Research*, *44*, 501–521.

Harmon-Jones, E., & Allen, J. J. B. (2001). The role of affect in the mere exposure effect: Evidence from psychophysiological and individual differences approaches. *Personality and Social Psychology Bulletin*, *27*, 889–898.

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

Hopf, J. M., Luck, S. J., Girelli, M., Hagner, T., Mangun, G. R., Scheich, H., et al. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex*, *10*, 1233–1241.

Hosokawa, T., Kato, K., Inoue, M., & Mikami, A. (2007). Neurons in the macaque orbitofrontal cortex code relative preference of both rewarding and aversive outcomes. *Neuroscience Research*, *57*, 434–445.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Kastner, S., Pinsk, M. A., De, W. P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*, 751–761.

Kim, Y. H., Gitelman, D. R., Nobre, A. C., Parrish, T. B., LaBar, K. S., & Mesulam, M. M. (1999). The large-scale neural network for spatial attention displays multifunctional overlap but differential asymmetry. *Neuroimage*, *9*, 269–277.

Kincade, J. M., Abrams, R. A., Astafiev, S. V., Shulman, G. L., & Corbetta, M. (2005). An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. *Journal of Neuroscience*, *25*, 4593–4604.

Kiss, M., Goolsby, B., Raymond, J. E., Shapiro, K. L., Silvert, L., Nobre, A. C., et al. (2007). Efficient attentional selection predicts distractor devaluation: ERP evidence for a direct link between attention and emotion. *Journal of Cognitive Neuroscience*, *19*, 1316–1322.

Korsten, N. J., Fragopanagos, N., Hartley, M., Taylor, N., & Taylor, J. G. (2006). Attention as a controller. *Neural Network*, *19*, 1408–1421.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*, 557–558.

Lepsien, J., & Nobre, A. C. (2007). Attentional modulation of object representations in working memory. *Cerebral Cortex*, *17*, 2072–2083.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1000–1014.

Nobre, A. C., Sebestyen, G. N., Gitelman, D. R., Mesulam, M. M., Frackowiak, R. S., & Frith, C. D. (1997). Functional localization of the system for visuospatial attention using positron emission tomography. *Brain*, *120*(Pt 3), 515–533.

Nunez, P. L., Wingeier, B. M., & Silberstein, R. B. (2001). Spatial-temporal structures of human alpha rhythms: Theory, microcurrent sources, multiscale measurements, and global binding of local networks. *Human Brain Mapping*, *13*, 125–164.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, *4*, 95–102.

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, *439*, 865–870.

Raymond, J. E., Fenske, M. J., & Tavassoli, N. T. (2003). Selective attention determines emotional responses to novel visual stimuli. *Psychological Science*, *14*, 537–542.

Raymond, J. E., Fenske, M. J., & Westoby, N. (2005). Emotional devaluation of distracting patterns and faces: A consequence of attentional inhibition during visual search? *Journal of Experimental Psychology: Human Perception Performance*, *31*, 1404–1415.

Sandberg, A., Tegner, J., & Lansner, A. (2003). A working memory model based on fast Hebbian learning. *Network*, *14*, 789–802.

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1999). Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *Journal of Neuroscience*, *19*, 1876–1884.

Seamon, J. G., Brody, N., & Kauff, D. M. (1983). Affective discrimination of stimuli that are not recognized: Effects of shadowing, masking, and cerebral laterality. *Journal of Experimental Psychology: Learning Memory and Cognition*, *9*, 544–555.

Stern, E. R., & Mangels, J. A. (2006). An electrophysiological investigation of preparatory attentional control in a spatial Stroop task. *Journal of Cognitive Neuroscience*, *18*, 1004–1017.

Sylvester, C. M., d'Avossa, G., & Corbetta, M. (2006). Models of human visual attention should consider trial-by-trial variability in preparatory neural signals. *Neural Network*, *19*, 1447–1449.

Taylor, J. G. (2003). Paying attention to consciousness. *Progress in Neurobiology*, *71*, 305–335.

Taylor, J. G., & Fragopanagos, N. F. (2005). The interaction of attention and emotion. *Neural Network*, *18*, 353–369.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670–674.

Tsotsos, J. K., Culhane, S. M., Wai, W., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*, 507–545.

Usher, M., & Niebur, E. (1996). Modelling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience*, *8*, 311–327.

Vuilleumier, P., & Schwartz, S. (2001). Emotional facial expressions capture attention. *Neurology*, *56*, 153–158.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Reviews*, *104*, 90–122.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, *5*, 277–283.

Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron*, *23*(4), 747–764.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, *1*, 202–238.

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

Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., et al. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature Neuroscience*, *5*, 995–1002.

Zucker, R. S., & Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual Review of Physiology*, *64*, 355–405.