eNeuro Accepted Manuscript

eNeuro

Research Article: New Research | Cognition and Behavior

Threat anticipation in pulvinar and in superficial layers of primary visual cortex (V1). Evidence from layer-specific ultra-high field 7T fMRI

https://doi.org/10.1523/ENEURO.0429-19.2019

Cite as: eNeuro 2019; 10.1523/ENEURO.0429-19.2019

Received: 14 October 2019
Accepted: 17 October 2019

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.eneuro.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Copyright © 2019 Koizumi et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

eNeuro Accepted Manuscript

Threat anticipation in pulvinar and in superficial layers of primary visual cortex (V1). Evidence from layer-specific ultra-high field 7T fMRI.

Ai Koizumi[1,2,3,4]*, Minye Zhan[5], Hiroshi Ban[1,2], Ikuhiro Kida[1,2], Federico De Martino[5], Maarten J. Vaessen[5], Beatrice de Gelder[5, 6]*, Kaoru Amano[1,2]

1. Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology (NICT)
2. Graduate School of Frontier Biosciences, Osaka University
3. Sony Computer Science Laboratories, Inc.
4. Graduate School of Media and Governance, Keio University
5. Department of Cognitive Neuroscience, Maastricht University
6. Department of Computer Science, UCL

Author Contributions: AK, HB, BDG, and KA contributed to the study design, FDM and IK contributed to the scanning and parameter settings of 7T fMRI, AK and MZ collected the data, AK conducted the data analyses with guidance and feedback from FDM, MZ, HB, MV and KA. AK, KA, and BDG wrote the manuscript with feedback from HB, MZ, FDM, and IK.

Correspondence should be addressed to: Ai Koizumi (bellkoizumi@gmail.com) & Beatrice de Gelder (b.degelder@maastrichtuniversity.nl)

Number of Figures: 4, Number of Tables: 1, Number of Multimedia: 0
Number of words for Abstract: 183, Number of words for Significance Statement: 103
Number of words for Introduction: 749, Number of words for Discussion: 1313

Acknowledgements
This research was funded by a European Research Council (ERC) grant under the European 512 Union Seventh Framework Programme for Research 2007–2013 (grant agreement number 513 295673), European Union's Horizon 2020 Research and Innovation Programme under grant 514 agreement No 645553, ICT DANCE (IA, 2015-2017), Incentive grant of National Institute of Information and Communications Technology (NICT), Japan, JSPS KAKENHI grant number 16K21687 and 18H02714 awarded to AK, and ERATO (JPMJER1801) awarded to HB. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Prof. Yuka Sasaki, Prof. Ben Seymour, Dr. Kazuhisa Shibata, Prof. Nim Tottenham, Dr. Brian Maniscalco, and Dr. Jorge Morales for their useful comments.

Conflict of Interest: No. Authors report no conflict of interest

Funding sources: Please find information in Acknowledgement.

eNeuro Accepted Manuscript

Abstract

The perceptual system gives priority to threat-relevant signals with survival value. In addition to the processing initiated by sensory inputs of threat signals, prioritization of threat signals may also include processes related to threat anticipation. These neural mechanisms remain largely unknown. Using ultra-high-field 7T fMRI, we show that anticipatory processing takes place in the early stages of visual processing, specifically in the pulvinar and V1. When anticipation of a threat-relevant fearful face target triggered false perception of not-presented target, there was enhanced activity in the pulvinar as well as in the V1 superficial-cortical-depth (layers 1–3). The anticipatory activity was absent in the LGN or higher visual cortical areas (V2-4). The effect in V1 was specific to the perception of fearful face targets and did not generalize to happy face targets. A preliminary analysis showed that the connectivity between the pulvinar and V1 superficial-cortical-depth was enhanced during false perception of threat, indicating that the pulvinar and V1 may interact in preparation of anticipated threat. The anticipatory processing supported by the pulvinar and V1 may play an important role in non-sensory-input driven anxiety states.

Keywords: threat perception; fearful face; 7 Tesla fMRI; V1 cortical layer, pulvinar

Significant statement

States of anxiety may typically be triggered by mere anticipation of threatening sensory inputs even without actual presentation. However, the mechanisms which mere anticipation of threat modulates the processing of incoming sensory inputs remains poorly understood, partly because the neural mechanisms underlying anxiety are often examined by measuring the effects initiated by visual presentation of actual threat stimuli. This study addresses how anticipation of threat modulates our visual system in its earliest stages. Specifically, this study shows that activity in the pulvinar and V1 is modulated based on anticipation of threat signals (fearful faces), leading to false perception of anticipated-yet-not presented threat signals.

eNeuro Accepted Manuscript

eNeuro Accepted Manuscript

Introduction

Fear and anxiety are core states of the organism and understanding these is a central issue for neuroscience. The importance of these psychological states is highlighted by findings that the visual system prioritizes inputs with high behavioural relevance (Bishop, 2007) such as threat signals from facial or bodily expression (Lanzetta and Orr, 1986; de Gelder, 2006; Koizumi et al., 2016). Fear is triggered by actual threat signals, but threat stimuli can also be anticipated and induce fear as is typically the case in anxiety (Grillon et al., 2017; Torrisi et al., 2018). The neural mechanisms underlying such anticipatory processing of threat, however, remains largely unknown.

We here hypothesize that when threat is anticipated, activity changes are seen in the early stage of visual processing in the pulvinar and V1. The literature postulates the pulvinar as a central relay, forwarding threat-relevant sensory inputs to other cortical and subcortical areas for quick evaluation and response (LeDoux, 1996; Pessoa and Adolphs, 2010; Tamietto and de Gelder, 2010; McFadyen et al., 2017). In addition to input-driven processing, the pulvinar is also known to receive inputs from higher cortex including prefrontal areas (Grieve et al., 2000; Bridge et al., 2016) presumably contributing to higher-level perceptual processing (Saalmann and Kastner, 2011; Kanai et al., 2015; Roth et al., 2016). In addition, we hypothesize that anticipatory processing of threat may involve the earliest stage of visual cortical hierarchy, V1. Studies using simple visual stimuli such as gratings have shown that V1 activity reflects not only the physically presented stimuli but also the subjective perception as in the case of false alarm (Ress and Heeger, 2003; Pajani et al., 2015). Relatedly, a recent study has demonstrated that visual images that are not physically present, but are well expected from the surrounding scenes, can induce the expected-image-like activity in V1 (Muckli et al., 2015). This effect was observed specifically in its superficial layers known to be modulated by the pulvinar in non-human primates (Shipp, 2003, 2007). These results suggest an active role of V1, especially its superficial cortical depth, in shaping visual perception in a top-down manner, at least when guided by some sensory inputs (Muckli et al., 2015).

To examine the role of the pulvinar and V1 in threat anticipation we designed an experiment where participants performed a simple task to detect a fearful face target (**Fig. 1**), which served as a social threat signal (Lanzetta and Orr, 1986; de Gelder, 2006). In a separate control session, a happy face detection task allowed us to examine whether the pulvinar and/or V1 contribute specifically to

eNeuro Accepted Manuscript

anticipation of threat or generalizes to any salient target. Importantly, participants were informed which target would be presented prior to each session in order to build their anticipation of a given target (Pajani et al., 2015). Prior knowledge combined with weakened sensory input due to brief presentation, induced participants' anticipation leading to more false percepts (Friston, 2005; Pajani et al., 2015). In each session, half of the trials randomly presented target faces (fearful or happy) and half neutral faces. Participants falsely perceived either a fearful or happy face target when a neutral face was actually presented in approximately 25% of the cases. We hypothesize that, unlike the trials where participants correctly perceive the presented fearful faces (i.e., HIT trials), the percept of a fearful or happy face in trials where a neutral face is presented (i.e., False Alarm, FA trials) cannot be explained by sensory input of the presented face but is instead likely caused by anticipatory top-down processing. Although it is expected that participants' anticipation for an upcoming threat-relevant fearful face target is noisy and fluctuates similarly across all trials, the proportion of trials where a higher level of anticipation contributes to the percept of a fearful face is likely to be larger among the FA than the HIT trials (**Fig. 1-1**).

We used ultra-high-field 7T fMRI with a spatial resolution of 0.8 mm to assess the activity of the pulvinar as well as V1 during the detection tasks. Such high spatial resolution enabled us to infer the cortical-depth dependent activity of V1 (Norris and Polimeni, 2019). The cortical-depth measurement is advantageous because the activity related to threat anticipation may be particularly observed in V1 superficial cortical depth, where non-sensory driven activity has been observed (Muckli et al., 2015). We predicted that the anticipation-driven false perception of a fearful face may be accompanied with enhanced activity in the pulvinar and V1, as well as the potentially enhanced functional connectivity between these two areas.

Materials and Methods

Participants

We enrolled 12 participants (6 males, mean age 23.73 ± SD 3.64, 2 left-handed), who provided written consent and received monetary reward after the experiment. Participants were all healthy and had normal or corrected-to-normal vision. The experiment protocol was approved by the ethical committee of

eNeuro Accepted Manuscript

Maastricht University. The data of one participant were removed from analysis due to excessive head motion (mean across runs > 4 mm).

We estimated that a sample size of N = 11 would be satisfactory to detect a medium to large effect size (f = .30) with an alpha power of 0.05 and power of 80% (G*Power version 3.1.9.2) as estimated with our 3T fMRI pilot results for pulvinar activity. Although we initially aimed for N = 12 to be conservative, one participant was removed from analysis as described above, leaving us with N = 11. Note that the sample size here is equivalent or larger than the related recent studies with 7T fMRI (e.g., four or 10 analyzed participants (Muckli et al., 2015; Kok et al., 2016).

Stimuli

Face images of 6 models (3 males) displaying fearful, happy, and neutral expressions were taken from the NimStim face stimulus set (Tottenham et al., 2009). We only included face images with the mouth open so that the local feature of an opened mouth alone would not enable detection of a fearful or happy rather than a neutral face. The images were grey-scaled and cropped into oval shapes to eliminate hair. They were then matched for luminosity, contrast, and spatial frequency spectrum with the SHINE toolbox (Willenbockel et al., 2010) implemented in MATLAB (R2011b, The Mathworks, Natick, Massachusetts, USA). We refrained from further manipulation of stimulus properties, as excessive manipulation itself could unintentionally induce differential activity in V1, which is sensitive to lower-level stimulus properties.

Experimental Design

Participants completed 2 sessions, in which either fearful or happy face targets were presented in a counterbalanced order.

The fearful face detection task required participants to detect a briefly presented fearful face followed by a mask (neutral face) (**Fig. 1**). On each trial, either a fearful or neutral face appeared as a target for 33 ms. Following a blank of 16.7 ms, a neutral face was presented for 133 ms as a mask that rendered the target face less visible but did not completely abolish its visibility. The model for the mask face was always different from the model for the target face. We used a neutral face as a mask instead of simpler images such as checkerboards, because V1 typically shows preferential activity towards simpler

eNeuro Accepted Manuscript

visual features (e.g., contrast and edges) and use of simple image masks could have interfered with the measurement of the critical activity in V1 related to face processing.

To further control task difficulty, the contrast of the target face was reduced to 35% of the contrast of the mask face, as determined during our pilot study. After 8.13 s after the target onset, response key assignment was shown on the screen. A fearful or neutral face response was assigned randomly to either the left or right key. Participants were instructed to respond with their right hand their first guess on whether they had perceived a fearful or neutral face target within a 2s time window. After a jittered inter-trial-interval from the onset of response assignment (8, 10.5, or 13 s), another trial was initiated. There were 24 trials in each of 8 runs (8 min 10 s per run), comprising 12 trials each for fearful and neutral face targets. There was no feedback provided on each trial. The order of trials was randomized.

In the control task session, we used happy faces, which are emotionally salient but non-threatening. The task was otherwise identical to the fear condition and used the same neutral face stimuli. The fearful and happy face sessions were conducted on two separate days in a counterbalanced order across participants. Stimuli were presented with Psychtoolbox (Brainard, 1997) implemented in Matlab (R2012a, the Mathworks).

For the data analysis the percept of participants was classified as follows: The correct percept of the presented fearful or happy target face was classified as HIT, whereas the false percept of a fearful or happy face in trials where a neutral face was presented was classified as FA. The correct percept of the presented neutral face was classified as Correct Rejection (CR), whereas the incorrect percept of a neutral face in trials where a fearful or happy face was presented was classified as MISS.

eNeuro Accepted Manuscript

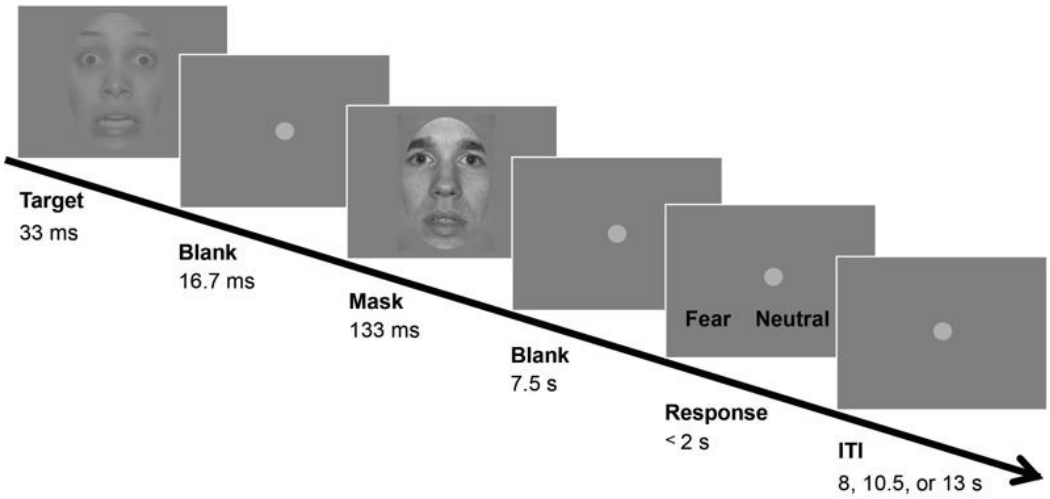

Figure 1. Design of the fearful face detection task. In each trial, either a fearful face target or a neutral face was presented briefly, followed by a mask consisting of a neutral face with a different identity than the target face. Participants responded whether they perceived a fearful target or neutral face by pressing the response key, which was randomly assigned trial-wise. The control task used happy face targets and neutral faces (not shown), and otherwise identical procedures. ITI: inter-trial interval. See also **Fig. 1-1.**

fMRI data acquisition

MRI data were acquired with a 7T Magnetom scanner (Siemens, Erlangen, Germany) at the Scannexus facility located at the Department of Cognitive Neuroscience, Faculty of Psychology and Neurosciences, Maastricht University (NL), with a Nova 1-transmitter/32-receiver head coil (Nova Medical, Wilmington, USA). For functional data acquisition, 2D gradient-echo planar images (EPI) were acquired at 0.8 mm isotropic resolution, with the following parameters: repetition time (TR) = 2500 ms, echo time (TE) = 21.8 ms, flip angle = 80, GRAPPA acceleration factor = 3, matrix size = 154 × 236, field of view (FOV) = 123 mm × 188 mm, slice thickness = 0.8 mm, number of slices = 40, no gaps, echo spacing = 1.04 ms, Bandwidth = 1116 Hz/Px, no multi-band acceleration. The slices were oriented to cover both the pulvinar and V1. To achieve maximal brain coverage with these parameters, right to left (RL) phase encoding was used for the task runs, so that the temporal areas outside the FOV were folded within the FOV and were trimmed later offline. A run of 5 TRs with the same parameters but with the opposing left to right (LR)

eNeuro Accepted Manuscript

phase encoding direction was acquired immediately before each task run for offline top-up EPI distortion correction (see *fMRI processing* for more details).

A separate run (3 min 30 s) was acquired to define V1 (see *Retinotopic delineation of V1*).

For anatomical data in 9 participants, a T1-weighted scan and a proton-density-weighted scan were acquired with a resolution of 0.6 mm isotropic (FOV = 229 mm × 229 mm, matrix size = 384 × 384, flip angle = 5. T1-weighted: TR = 3100 ms, TE = 2.52 ms; proton-density-weighted: TR = 1440 ms, TE = 2.52 ms). For the other 3 participants, we used anatomical images with a spatial resolution of 0.7 mm isotropic from previous unrelated experiments.

fMRI processing

fMRI analyses were conducted in BrainVoyager 20.2 (Brain Innovation, Maastricht, the Netherlands). For preprocessing of fMRI data, we trimmed the lateral sides of EPI images by a small amount (60 voxels) to remove the folded-in tissue outside the FOV. The folding-in and trimming did not affect the coverage of the bilateral pulvinar and V1. The trimmed EPI images were then slice time corrected (sinc interpolation) and corrected for 3D rigid body motion (trilinear/sinc interpolation). Distortions of the EPI images from the task runs were adjusted against EPI images taken immediately before each task run with the opposing encoding phase (Andersson et al., 2003), with the BrainVoyager plugin COPE (http://support.brainvoyager.com/available-tools/49-available-plugins/477-cope-plugin-for-epi-distortion-correction.html). EPI images then underwent temporal high-pass filtering with 2 cycles per run. That is, signals with temporal frequency lower than half the length of a run (i.e., 250 s) were removed with discrete Fourier filter.

After the preprocessing EPI images were manually aligned to the anatomical images in BrainVoyager including optimization of alignment around the posterior portion of the brain encompassing V1 and the pulvinar. From the total of runs those with 3D motion larger than 2.5 mm were discarded from analyses, because large motion induced excessive and/or unique EPI distortion that interfered with precise alignment and the subsequent cortical depth specific analyses. One participant's data were excluded from further analysis due to excessive head movements (> 4 mm). For the remaining participants, the run numbers included in analyses did not differ between the fearful face task (M = 7.64 ±

eNeuro Accepted Manuscript

SD 0.67) and the happy face task (M = 7.45 ± 0.82) ($t(10) = 0.482$, $p = 0.640$). The mean numbers of trials entered into the fMRI analyses for each participant, after excluding runs with excessive head motion were: fearful HIT, M = 45.64 ± s.e. 4.0; fearful FA, 26.64 ± 3.48; happy HIT, 54.18 ± 4.32; happy FA, 21.18 ± 2.93.

Task-related activity was then estimated with a deconvolution analysis, in which responses for successive 5 points (2.5 s × 5 TRs) were estimated for each of the 8 conditions, starting from the onset of the target face for each of the 8 experimental conditions (fearful face HIT, MISS, FA, CR, as well as happy face HIT, MISS, FA, CR). We used the deconvolution analysis because it does not assume a fixed haemodynamic response function, which is generally built based on the response properties in sensory cortical areas (Boynton et al., 1996; Glover, 1999) and is thus potentially less favorable for thalamic areas. The entire deconvolved time course in V1 and the pulvinar are shown in **Fig. 2-1**. Additionally, the GLM model included 6 head-motion nuisance parameters (3 translation directions and 3 rotation axes). EPIs were spatially smoothed only when localizing the task-relevant voxels but were not smoothed when estimating the task-related activity to maintain laminar specificity.

Localization of the pulvinar and LGN

We localized the pulvinar as the target region, as well as the LGN as a control region (**Fig. 2**). Each of the thalamic regions (the pulvinar and LGN) was first anatomically defined based on their physical properties. Specifically, the pulvinar was located in the dorsal thalamus that is superior and medial to the LGN, located adjacent to the third ventricle (Kastner et al., 2004; Sprenger et al., 2012; Mai, 2015). To localize the pulvinar, we referred to the histological atlas (Chakravarty et al., 2006) superimposed on the high-resolution T1 resampled at 0.8 mm in native space. In order to specifically select task-relevant voxels within the anatomically localized pulvinar, functional images from the task runs were smoothed with a gaussian kernel of 2.4 mm full width at half maximum (FWHM). Task-relevant voxels were then defined based on activation at the temporal peak of the time course (5 s from the target face onset while considering a haemodynamic delay of 5 s, i.e., 2 TRs × 2.5 s) that was larger than baseline at a threshold of $p < 0.01$ uncorrected to compensate for a generally lower tSNR in subcortical areas (see *Estimation of*

eNeuro Accepted Manuscript

tSNRs). Note that task-relevant voxels were selected based on the target face onsets from all trials including all conditions, to minimize any bias towards one particular condition over another.

Pulvinar ROIs were located in both hemispheres for all participants except for 1 participant (ROI in the left hemisphere only). LGN ROIs were located in both hemispheres for 5 participants, while they were located in either the right or left hemisphere for 4 and 2 participants, respectively. Inability to locate the ROIs in both hemispheres in some participants is likely to be due to the generally hindered SNR towards deeper brain areas (see *Estimation of tSNRs*). For participants with pulvinar and/or LGN ROI(s) located in both hemispheres, activity was estimated for each hemisphere and averaged.

Retinotopic delineation of V1

To delineate V1 in each participant, a retinotopy run was acquired. During the run, color/luminance-flickering wedge-shaped checkerboard patterns (30 deg in polar angle) were presented along the horizontal or vertical meridian alternately for 15 s, each with 6 repetitions, following procedure in (Lafer-Sousa and Conway, 2013). Checkerboard patterns were flickered at 4 Hz and were displayed in 1 of 4 color combinations (red/green, blue/yellow, black/white, and magenta/cyan) to activate neurons with various response profiles and enhance the signals to identify the boundaries between the cortical areas. The boundaries of V1 were delineated with a general linear model contrasting activity between the horizontal and vertical presentation periods as previously described (Lafer-Sousa and Conway, 2013). V2, V3, and V4 were delineated in a similar manner and served as control regions. The boundary of V4 was located while additionally referring to its predefined anatomical landmarks (Witthoft et al., 2014).

Anatomical image processing and cortical-depth specific estimation of V1 activity

Inhomogeneity of T1-weighted images was corrected by dividing the original image intensities by the proton density images (Van de Moortele et al., 2009). Subsequently, the corrected T1 image was resampled at a resolution of 0.8 mm to match the resolution of EPI. The boundaries of grey-white matter and the pial surface were first estimated with BrainVoyager 20.2, and further corrected manually, to improve the precision and to remove the blood vessels and dura mater based on image intensity. The

eNeuro Accepted Manuscript

anatomical image was not transformed to standardized coordinates but was kept in native space, to reduce resampling and maintain its laminar properties undistorted.

As was done for the pulvinar, the task-relevant voxels were selected within the delineated V1 of each hemisphere. Specifically, the task-relevant voxels were defined based on the contrast between all target face onsets versus baseline at a threshold of $p < 0.001$ uncorrected, at the temporal peak of the time course at 5 s (2 TRs × 2.5 s). The task-relevant voxels within V1 were successfully located in both hemispheres in 8 participants, while they were located in only one hemisphere in the remaining 3 participants (right only, N = 1; left only, N = 2). For participants with peaks located in both hemispheres, the estimates of activity were averaged between the hemispheres for each visual cortical area (e.g., V1).

To define the cortical depths of V1, we used the Laplace equation to estimate cortical thickness and then obtained an equidistant definition of depth with respect to the local thickness (Muckli et al., 2015; De Martino et al., 2018) at three depth levels (from 25, 50, and 75% depth levels relative to the local cortical depth in an inward direction) centering around the spatial activity peak (i.e., voxel with highest activation level) among the pre-defined task-relevant voxels with 15 × 15 grids of 0.5 voxels. Individual voxels were assigned to the adjacent cortical depth, and were used as regions of interest (ROIs) in the subsequent analyses (**Fig. 3**). We confined our analyses to voxels allocated to the superficial and deep cortical depth groups, which roughly correspond to cortical layers 1 to 3 and layers 5 to 6, respectively, due to the difference in anatomical thickness of each layer (de Sousa et al., 2010; Kok et al., 2016). The voxels assigned to the superficial and deep depths covered an average of 22.67% (s.e. = 3.95) of the retinotopically delineated V1 (see *V1 delineation*). This relatively small coverage by the task-relevant voxels is likely to be due to lower contrast and luminosity as well as smaller stimulus size relative to the V1 delineation run (i.e., The visual angles spanned by the checkerboards wedges and the task face images were 10.4° × 10.4° and 6.1° × 4.5°, respectively). The cortical-depth dependent activity was estimated from the unsmoothed functional data in order to maintain the original spatial resolution (0.8mm).

We localized the control areas V2-4 in a similar manner as V1. The task-relevant voxels within V2 and V3 were successfully located in both hemispheres in 9 participants, while they were localized in only one hemisphere in 2 participants (V2 right only, N = 1; V2 left only, N = 1; V3 right only, N = 1; V3 left

eNeuro Accepted Manuscript

only, N = 1). The cortical depths of V2 and V3 were defined separately for the dorsal and ventral areas, and the voxels for each cortical depth were combined between the dorsal and ventral areas. The task-relevant voxels within V4 were located in both hemispheres in all participants.

Estimation of tSNRs

The SNR of the time series (tSNRs) for the pulvinar, LGN, and V1 superficial and deep cortical depths were assessed with VTC inspector plugin in BrainVoyager (Brain Innovation). tSNR was assessed for each ROI from the time course in initial runs of the tasks and averaged across participants. The mean tSNR (± s.d.) was 8.12 (± 1.03) for the pulvinar, 7.06 (± 1.38) for LGN, 18.15 (± 4.48) for V1 superficial cortical depth and 18.04 (± 3.92) for V1 deep cortical depth. In line with the previous report that the gradient-echo imaging sequence employed here yields higher SNR towards the surface of the cortex (De Martino et al., 2018), thalamic areas (i.e., the pulvinar and LGN) that were deeper inside the brain and further from the coil had lower tSNR relative to V1 on the outer brain.

Generalized form of context-dependent psychophysiological interaction analysis (gPPI)

To examine whether functional connectivity between the pulvinar and V1 was enhanced during certain trial types (e.g., FA trials with fearful faces), we conducted a pair of gPPI analyses (McLaren et al., 2012): one with V1 superficial cortical depth voxels as the seed ROI, and the other with V1 deep cortical depth voxels as the seed ROI. For each analysis, the GLM model included regressors for each of the eight trial types (i.e., HIT, FA, MISS, CR for fearful and happy faces) convolved with the canonical two-gamma HRF, a regressor for the z-normalized time course of the seed ROI, regressors for PPI terms (i.e., seed time course × trial type regressor), six nuisance regressors of 3D head motion (3 translation directions and 3 rotation axes). The GLM analysis was run for each participant, and the parameter estimates (Beta values) were extracted within the pulvinar ROI (see *Materials and Methods* for ROI definition) for the PPI term of HIT and FA trials of fearful and happy faces (i.e., the four critical trial types included in the main results). In the group-level analysis, t-values for the parameter estimates for each trial type and seed ROI in gPPI (**Fig. 3-2**) were tested against 0 with a one-sample t-test, with Bonferroni correction for eight conditions (i.e., p = 0.05/8, HIT/FA × Fearful/Happy × V1 superficial/deep cortical depths).

346

347 **Statistical Analysis**

348 We conducted subject-level analyses of fMRI data in BrainVoyager 20.2 (Brain Innovation, Maastricht,
349 The Netherlands) (see *fMRI processing*), and subsequently, conducted group-level repeated measures
350 ANOVAs in IBM SPSS Statistics (version 18). The assumption of sphericity was met in the current
351 dataset because there were only two levels per factor (e.g., emotion) in each repeated ANOVA.
352 Behavioural performance was also analyzed in IBM SPSS.

353 Following previous studies (Keselman and Keselman, 1993; Tamaki et al., 2016), when an
354 omnibus 3-way repeated measures ANOVA revealed a significant 3-way interaction, we conducted 2-way
355 (n-1 way) repeated measures ANOVAs to locate a simple interaction effect. A significant 2-way interaction
356 that was identified was subsequently tested with a t-test (two-tailed) to examine simple main effects.
357 Similarly, when an omnibus repeated measures ANOVA initially involved only two factors (i.e., 2-way), a
358 significant interaction was similarly followed by t-tests (two-tailed). The series of t-tests were not
359 susceptible to the inflation of type 1 error as they followed significant interactions in the initial omnibus
360 ANOVAs, as has been validated and commonly practiced previously (Keselman and Keselman, 1993;
361 Cohen, 2004). We estimated effect sizes in the post-hoc t-tests with Cohen's d, where a value larger than
362 0.8 indicated a large effect size and a value larger than 0.5 indicated a medium effect size (Cohen, 1988;
363 Tamaki et al., 2016). All t-tests reported in this manuscript were two-tailed.

eNeuro Accepted Manuscript

eNeuro Accepted Manuscript

Results

Behavioural performance

Perception of a fearful face reported on presentation of a fearful face target was classified as HIT, whereas perception of a fearful face reported on presentation of a neutral face was classified as a false alarm (FA) and similarly for the happy vs. neutral faces in the control task. The ratio of FA was similar between the fearful face and the happy face detection task (M = 28.51 ± 3.52%; M = 24.13 ± 3.07%, respectively), and the rates in each task did not significantly differ from each other ($t(10) = 1.845$, $p = 0.095$[a]). The ratio of HIT trials was higher for the happy face (M = 60.36 ± s.e. 4.36%) than for the fearful face detection task (M = 49.18 ± 4.42%) ($t(10) = -4.155$, $p = 0.002$[b]). This is consistent with previous literature showing that explicit labelling of happy faces is easier than that of negative faces (Calvo and Lundqvist, 2008). Our choice of facial images with open mouth may have additionally contributed to relatively poorer detection of a fearful face against a neutral face target or mask because a neutral face may look more similar to a fearful face with opened mouth.

Pulvinar activity during the detection tasks

Our primary interest was how activity of the higher-order thalamic area pulvinar (**Fig. 2A**) may differ between HIT and FA trials, and whether such a difference is specific to the fearful face detection task.

We observed that pulvinar activity was greater on FA trials than on HIT trials (**Fig. 2B**), as demonstrated by a main effect of percept type in a repeated measures ANOVA ($F(1,10) = 8.310$, $p = 0.016$). Greater activity in FA trials relative to HIT trials was observed during the fearful face detection task ($t(10) = -2.944$, $p = 0.015$, $d = 0.8$[c]) but not during the happy face detection task ($t(10) = -1.23$, $p = 0.247$, n.s.[d]). There was no significant difference between FA trials of the fearful face detection task and that of the happy face detection tasks ($t(10) = .572$, $p = .580$, n.s.[e]). There was no significant interaction between percept type and emotion ($F(1,10) = 0.155$, $p = 0.702$). The main effect of percept type independent of emotion suggests that the pulvinar shows enhanced activity with a false percept in general, as has been reported in a previous study during a false detection of change in non-emotional stimuli (Pessoa and Ungerleider, 2004) (see also **Fig. 2-1** for the deconvolved time course of the pulvinar).

eNeuro Accepted Manuscript

As a control analysis for the pulvinar activity, we examined the activity of a first-order thalamic region, the LGN (**Fig. 2C**). Our analysis showed that the activity level of the LGN was not altered in FA trials of fearful faces (**Fig. 2D**). Neither the main effect of percept type nor that of emotion was significant in a repeated-measures ANOVA ($F(1,10) = 0.100$, $p = 0.759$; $F(1,10) = 3.014$, $p = 0.113$, respectively[f]). The two-way interaction was also non-significant ($F(1,10) = 0.363$, $p = 0.560$). The result of this control analysis suggests the pulvinar, rather than the thalamic areas in general, contributed to FA trials.

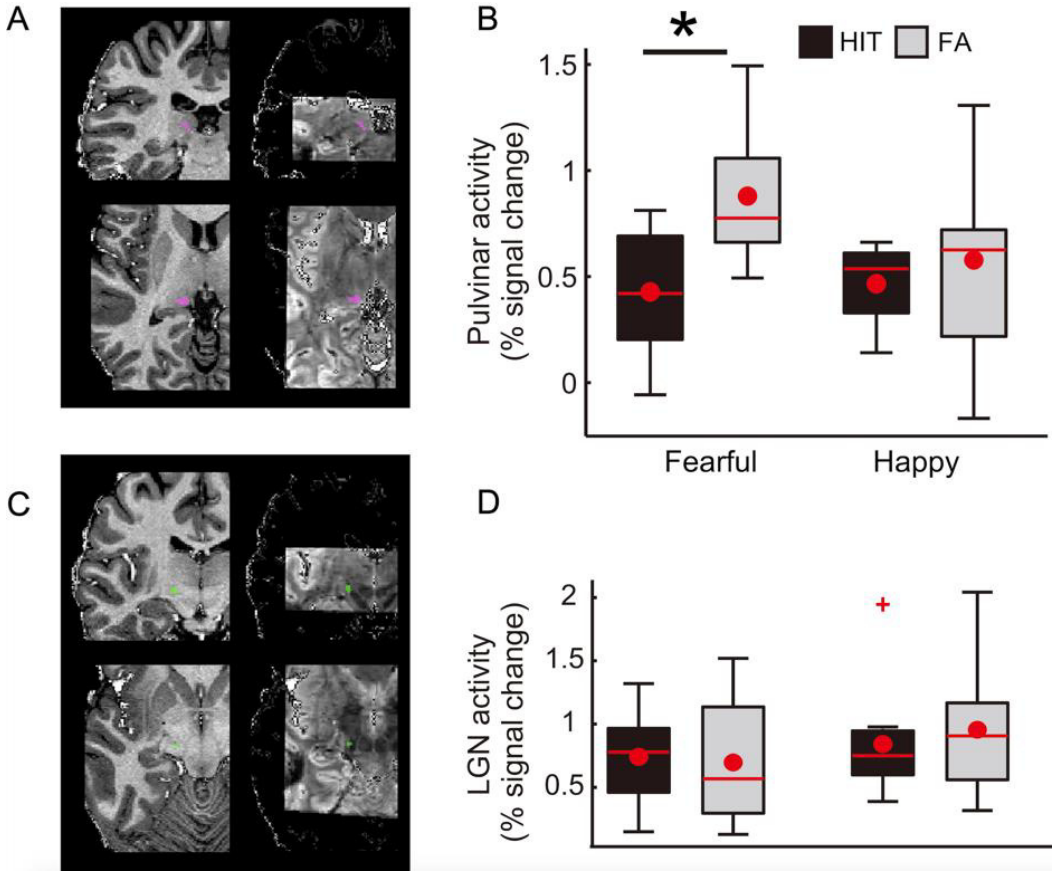

Figure 2. Activity in the pulvinar and LGN. A. Demonstrations of pulvinar ROIs from a representative participant (see *Materials and Methods* for ROI definition), shown on the anatomical image (left panels) and EPI (right panels). **B.** Pulvinar showed enhanced activity in FA trials relative to HIT trials during the fearful face detection task ($t(10) = -2.944$, $p = 0.015$, $d = 0.8$) but not during the happy face detection task

eNeuro Accepted Manuscript

(t(10) = -1.23, p = 0.247, n.s.). There was no significant interaction between percept type and emotion (F(1,10) = 0.155, p = 0.702). Here, data from one participant with an extreme outlier (activity exceeding 5 standard deviations above the group mean across four conditions, i.e., HIT/FA x fearful/happy) are not shown. **C.** Demonstrations of the LGN ROI from an example participant (see *Materials and Methods* for ROI definition). **D.** Unlike the pulvinar, the LGN showed no differential activity between the percept types and facial emotions. Box plot shows upper (75%) and lower (25%) quartiles with median (red line) and mean (red dot), with whisker showing maximum and minimum value. An outlier (outside of ± 2.7 standard deviations within a distribution for a given condition) is shown with a red cross. *p < 0.05. See also **Fig. 2-1 and 2-2.**

Cortical-depth dependent V1 activity during fearful face perception

While pulvinar activity alone did not dissociate between fearful and happy faces (**Fig. 2**), we observed that V1 superficial activity did (**Fig. 3D**). A repeated measures analysis of variance (ANOVA) revealed a second-order interaction between percept type (HIT/FA), emotion (fearful/happy), and cortical depth (superficial/deep) of V1 (F(1,10) = 5.667, p = 0.039). This interaction was due to the fact that percept type and emotion interactively modulated V1 activity at superficial cortical depths (F(1, 10) = 7.840, p = .019), but not at deep cortical depths (F(1,10) = 0.673, p = .431) (see **Fig. 2-1** for the deconvolved V1 time course and **Fig. 3-1** for the results separately plotted for individual participants).

At the superficial depth, V1 activity in FA trials of a happy face was less than that in HIT trials of a happy face (t(10) = 2.387, p = 0.038, d = 0.9[g]). This result is consistent with a previous study demonstrating that V1 activity levels are typically greater for HIT than for FA trials in a visual detection task with non-threatening targets (Ress and Heeger, 2003), although cortical depth dependent activities were not reported.

Contrary to the happy face detection task, V1 superficial depth activity in FA trials during the fearful face detection task was similar to or even numerically larger than that in HIT trials (t(10) = -1.714, p = 0.117, n.s.[h]). Indeed, post-hoc analysis showed that V1 superficial depth activity in FA trials of a fearful face was significantly greater than that in FA trials of a happy face (t(10) = -2.486, p = 0.032, d = 0.8[i]), even though the same neutral faces were presented with the only difference being the task context to

anticipate either a fearful or happy face target. These results are consistent with the idea that anticipation-driven activity, which is more evident in the FA trials than in the HIT trials, was added on top of the smaller input-driven activity in FA trials. In other words, our results suggest that V1 superficial depth activity may reflect excessive top-down processing in anticipation of fearful face targets, which does not generalize to mere anticipation of any emotionally salient target (i.e., happy faces).

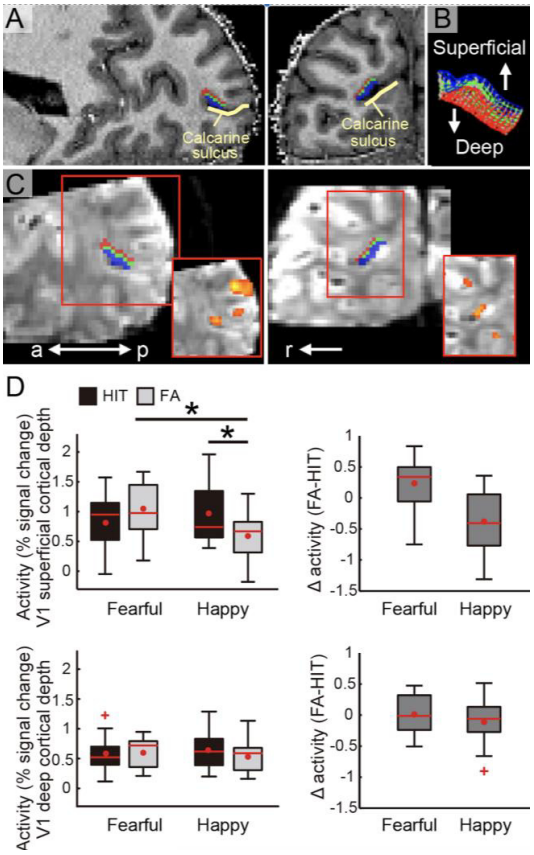

Figure 3. Demonstrations of V1 cortical depths and cortical-depth dependent activity during the tasks. A. V1 cortical depth is visualized on an anatomical image of a representative participant with sagittal (left panel) and coronal views (right panel). The voxels allocated to superficial (outwards to pial surface) and deep (inwards to white matter) cortical depths are shown in blue and red, respectively. The voxels allocated to the intermediate depth (shown in green) were disregarded in the main analyses (see *Materials and Methods*). **B.** The cortical grid mesh within which the voxels were allocated. **C.** The V1

eNeuro Accepted Manuscript

eNeuro Accepted Manuscript

cortical depth is visualized on EPI images (visualized in 3D for demonstrative purpose) in sagittal and coronal views (left and right, respectively). Red squares at the lower right demonstrate activity for all face targets relative to baseline on the EPI smoothed with a 3D kernel of 2.4 mm full width at half maximum (FWHM). **D.** Peak activity at V1 superficial and deep cortical depths (upper and lower rows, respectively) in HIT and false alarm (FA) trials in the fearful face and happy face detection tasks. The difference in activity between FA and HIT trials are demonstrated in the right panel for each task, with a larger value indicating greater activity for FA than for HIT trials. Box plot shows upper (75%) and lower (25%) quartiles with median (red line) and mean (red dot), with whisker showing maximum and minimum value. An outlier (outside of ± 2.7 standard deviations within a distribution for a given condition) is shown with a red cross. a: anterior, p: posterior, r: right, *$p < 0.05$. See also **Fig. 2-1, 3-1, 3-2, 3-3, 3-4.**

Connectivity between the pulvinar and V1 during fearful face perception

The aforementioned results showed that V1 superficial depth activity in FA trials was enhanced only for fearful faces and pulvinar activity was higher in FA trials than in HIT trials regardless of face type. Given that there is a reciprocal interaction between the pulvinar and V1 including modulatory input from the pulvinar to V1 superficial layers (Shipp, 2003; Purushothaman et al., 2012; Cicmil and Krug, 2015; Bridge et al., 2016; Roth et al., 2016; Shipp, 2016), it is possible that the enhanced V1 superficial depth activity in FA trials of fearful faces is at least partly related to its communication with the pulvinar. We examined whether a modulation of activity was present between the pulvinar, and the superficial or deep V1 voxels using a generalized form of context-dependent psychophysiological interactions analysis (gPPI) (McLaren et al., 2012). A functional coupling was present between the pulvinar and V1 superficial depth in FA trials for fearful faces ($t(10) = 3.981$, $p = 0.0026$, one sample t-test against 0, CI [0.319 1.128] [j]; **Fig. 3-2**), but not in the other conditions tested (fearful HIT: $t(10) = 0.403$, $p = 0.695$, CI [-0.683 0.985]; happy HIT: $t(10) = 1.869$, $p = 0.091$, CI [-0.105 1.120]; happy FA: $t(10) = 1.4364$, $p = 0.181$, CI [-0.271 1.253]). In addition, another analysis hinted that there was already enhanced activity in the pulvinar prior to the onset of FA trials with fearful faces but not in V1 superficial depth (**Fig. 3-3**). This result, together with the result of gPPI analysis, suggests that the modulatory input from the pulvinar may have contributed to the enhanced V1 superficial activity in FA trials with fearful face percept, although future studies need to

eNeuro Accepted Manuscript

directly test the specific direction of interaction between the two regions. There was no significant interaction with V1 cortical depth and facial emotions in the gPPI result ($F(1,10)=.198$, $p = .666$) (**Fig. 3-2**).

Control analyses

To examine the specificity of our results, we conducted a series of control analyses. We particularly examined the potential additional contributions of higher visual cortical areas, V2, V3, and V4 in the FA trials for fearful faces. The activity of these areas was examined separately for their superficial and deep cortical depths in a manner similarly to V1 (see **Fig. 4A** and *Materials and Methods*). Other higher visual areas such as fusiform areas were outside of the fMRI coverage.

Unlike V1, we found that the activity of V2-V4 was not enhanced in the FA trials of fearful faces (**Fig. 4B**). With V2, a second-order interaction between percept type, emotion, and cortical depth was non-significant ($F(1,10) = 0.397$, $p =0.543$). The main effects of percept type and emotion were also non-significant ($F(1,10) = 0.612$, $p = 0.452$; $F(1,10) = 0.240$, $p = 0.635$, respectively), while only the main effect of cortical depth was significant ($F(1,10) = 21.245$, $p = 0.001$). Similarly, with V3, a second-order interaction was non-significant ($F(1,10) = 0.265$, $p = 0.618$). Although there was a general trend that the activity on HIT trials was larger relative to FA trials, the main effect of percept type did not reach significance ($F(1,10) = 4.534$, $p = 0.059$). The main effect of percept type was also non-significant ($F(1,10) = 0.018$, $p = 0.895$), while a main effect of cortical depth was significant ($F(1,10) = 20.434$, $p = 0.001$). Likewise, with V4, a second-order interaction ($F(1,10) = 1.028$, $p = 0.335$) as well as the main effects of percept type and emotion were non-significant ($F(1,10) = 0.124$, $p = 0.732$; $F(1,10) = 0.850$, $p = 0.37$, respectively), while only the main effect of cortical depth was significant ($F(1,10) = 18.523$, $p = 0.002$).

As an additional control analysis, we examined whether similar results to those observed in FA trials of fearful faces would be present in MISS trials, in which a fearful face was presented but not detected. While the aforementioned results for FA trials of fearful faces may reflect enhanced top-down processing as we speculated earlier, other non-mutually exclusive possibilities are worth considering. Specifically, the results in FA trials may reflect a mere ‘missing’ of a presented face image, in which case

eNeuro Accepted Manuscript

similar results as FA trials would be expected in MISS trials. Contrary to this possibility, we did not observe any notable results specific to MISS trials of fearful faces either in V1 or in the pulvinar (a non-significant interaction between percept type (MISS/CR) and emotion: $F(1,10) = 0.013$, $p = 0.912$) (**Fig. 3-4**), excluding the possibility that the results in FA trials of fearful faces merely reflected a mere missing of a presented face image yielding a mismatch between the input and percept.

Taken together, the analyses suggest that the results of pulvinar and V1 superficial cortical depth in false perception of fearful faces were not mirrored in the LGN or in V2-4. The findings also indicate that the results of pulvinar and V1 were not explained away either by mere anticipation of any emotionally salient target (i.e., happy face) or by mere mismatch between percept and input, suggesting the specificity of the involvement of pulvinar-to-V1 input in false perception of fearful faces.

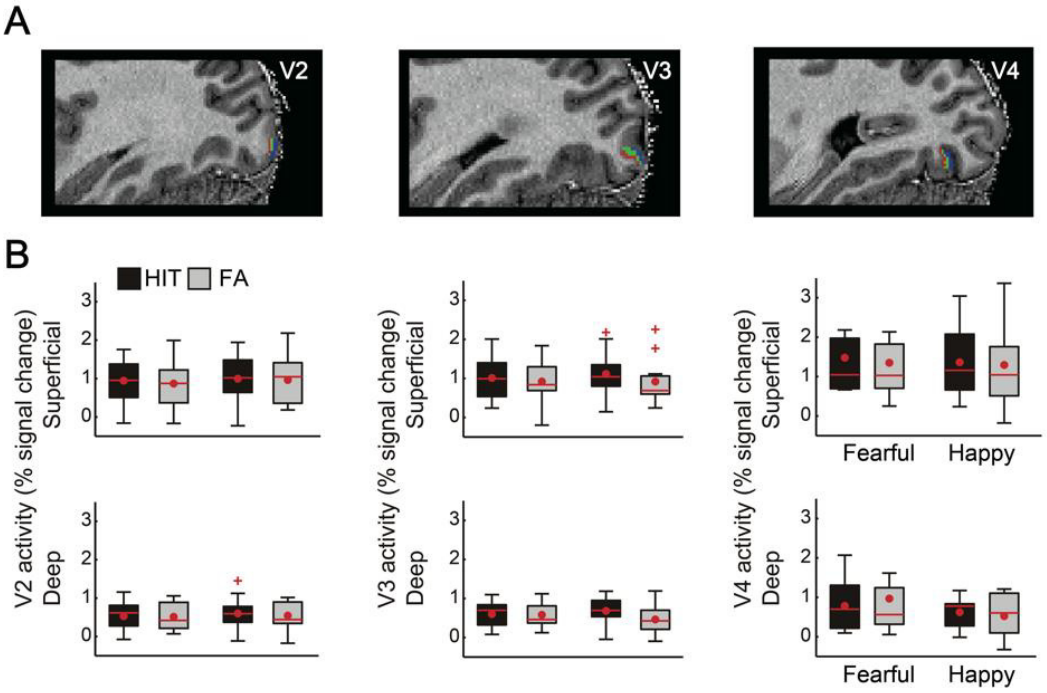

Figure 4. Control analyses showing no differential activity in V2, V3, and V4 in HIT compared with false alarm (FA) trials. A. V2, V3, and V4 cortical depths are visualized on the anatomical image of an example participant, with the same color coding as for V1 shown in **Fig. 3** (see *Materials and Methods* for

ROI definition). **B.** Unlike V1, V2-4 did not show any differential activity between the percept types and facial emotions, regardless of cortical depth. Box plot shows upper (75%) and lower (25%) quartiles with median (red line) and mean (red dot), with whisker showing maximum and minimum value. An outlier (outside of ± 2.7 standard deviations within a distribution for a given condition) is shown with a red cross.

eNeuro Accepted Manuscript

Statistical Table

	Data structure	Type of test	Confidence Interval
a	Normal distribution	Paired-sample t-test (two-tailed)	[-0.173 -0.052]
b	Normal distribution	Paired-sample t-test (two-tailed)	[-0.009 0.097]
c	Normal distribution	Post-hoc t-test following ANOVA (two-tailed)	[-0.782 -0.108]
d	Normal distribution	Post-hoc t-test following ANOVA (two-tailed)	[-0.888 0.256]
e	Normal distribution	Post-hoc t-test following ANOVA (two-tailed)	[-0.391 0.661]
f	Normal distribution	Post-hoc t-test following ANOVA (two-tailed)	fearful [-0.335 0.424]; happy [-0.500 .272]
g	Normal distribution	Post-hoc t-test following ANOVA (two-tailed)	[0.025 0.735]
h	Normal distribution	Post-hoc t-test following ANOVA (two-tailed)	[-0.542 0.071]
i	Normal distribution	Post-hoc t-test following ANOVA (two-tailed)	[-0.871 -0.048]
j	Normal distribution	One sample t-test (two-tailed)	[0.319 1.128]

Discussion

We observed that the false percept of anticipated but not present fearful face relative to that of a happy face was accompanied by increased activity in superficial cortical depth of V1, which constitutes the earliest stage of the visual cortical hierarchy. This enhanced V1 activity may particularly contribute to the anticipatory perception of threat signals, as it did not generalize to the perception of non-threatening signals, i.e., happy faces. Although preliminary, we additionally observed that the connectivity between the pulvinar and V1 superficial cortical depth was enhanced when participants falsely perceived fearful faces. These results suggest a potential role of the pulvinar and V1 in preparing the visual system to perceive an anticipated threat.

It may be counterintuitive to expect a crucial role of V1 in the perception of fearful faces, given that V1 is fine-tuned to low-level visual features that would constitute only low-level properties of fearful faces. Nevertheless, recent findings suggest that a lower visual cortical area can reflect higher-level features, when the prediction signals for such features originate from a higher area (Schwiedrzik and Freiwald, 2017). While such top-down signals are typically expected to descend from higher cortical areas, it has recently been speculated that the pulvinar also contributes to such top-down signals (Kanai et al., 2015), including contextual signals (Roth et al., 2016). Given that the pulvinar appears capable of

eNeuro Accepted Manuscript

eNeuro Accepted Manuscript

539 coding threat signals, including a complex fearful face (de Gelder, 2006; Pessoa and Adolphs, 2010), one
540 possibility is that the pulvinar serves as one of the critical regions to modulate the early stage processing
541 in V1 rapidly biasing visual cortical processing towards threat perception.

542 In line with this, a recent MEG study with human participants demonstrated that the neural activity
543 driven by the inputs of facial images is better explained by a dynamic causal model which considers the
544 pulvinar-to-V1 input, and that such input is modulated by the presence of fearful expression in faces
545 (McFadyen et al., 2017). While this study suggests the role of the pulvinar in input-driven relaying of
546 information to V1, our study suggests that the pulvinar may play an additional role in modulating activity in
547 the early stage of visual cortex in anticipation of threat-relevant signals in humans.

548 Limitations of the current study do not allow a more direct conclusion of the direction of the
549 interaction. First, the temporal resolution of fMRI was not high enough to detect the precedence of one
550 region over the other, which may be compensated for in a future study using MEG with higher temporal
551 resolution. Secondly, the portions of the pulvinar active for the FA of fearful faces were not always
552 symmetrical between the hemispheres and did not always converge to the anatomical subregions that
553 have direct communication with V1 (Bridge et al., 2016) (Figure 2-2). As the localization of the subregions
554 of the pulvinar may be partly precluded because of relatively lower SNR (see *Estimation of tSNRs*), future
555 studies may adopt alternative fMRI sequences for a better signal in the pulvinar. Future studies may also
556 examine whether the current results generalize when facial images and masks were controlled differently
557 (e.g., faces with closed mouth and non-facial masks).

558 The absence of enhanced V2-V4 activity in the FA trials of fearful faces may be surprising
559 considering that the pulvinar and V1 could both drive V2 activity (Marion et al., 2013). However, this null
560 result does not necessarily indicate that the enhanced activity in the pulvinar and/or V1 have no
561 subsequent effect in higher visual areas V2-V4 in the FA trials of fearful faces. This is because the input-
562 driven activity in V2-V4 is supposedly higher in HIT trials relative to FA trials (Ress and Heeger, 2003),
563 especially due to salient sensory inputs of the fearful face targets. Thus, the absence of activity difference
564 between the FA trials and HIT trials in these downstream areas of V1 may instead suggest that the
565 anticipation-driven activity in the pulvinar and/or V1 may have compensated the originally lower activity in
566 V2-V4 in the FA trials of fearful faces, making such activity comparable to the HIT trials of fearful faces.

eNeuro Accepted Manuscript

Note that these null results do not exclude that other areas besides the pulvinar and V1 were involved in false percept (i.e., FA trials) of fearful faces. Future studies should investigate whether and how the pulvinar and V1 may interact with other areas, such as the amygdala, fusiform face area and prefrontal areas not covered in this study. Whole brain analyses, which may help address such questions, were not performed because of the restricted brain coverage (see *fMRI Data acquisition* in *Materials and Methods*), and because the spatial distortion and/or SNR were expected to be highly inhomogeneous even across the covered brain areas due to ultra-high magnetic field (see *Estimation of tSNRs*).

How false perception of non-threatening cues, such as happy faces, emerges remains to be investigated. One possibility is that such a percept would reflect top-down modulation of visual cortical areas higher than V1, similarly to the processing of facial identity (Schwiedrzik and Freiwald, 2017). Although the control analyses did not support the contribution of V2-4 to the false perception of happy face targets, one possibility may be that visual areas higher than V4 contributed here, although these higher areas were not covered in this study (to allow for the spatial resolution desired).

How the subjective experience of falsely perceiving an unpresented fearful face may have differed from that of correctly perceiving a presented fearful face is one remaining interesting question. One way to examine the potential difference could be to assess perceptual confidence. Another way could be to measure the whole brain activity to examine whether other brain areas, e.g., motor areas as well as the pulvinar also implicated in action (Dominguez-Vargas et al., 2017), may have also contributed to bias responses in false perception, although it is noteworthy that response bias could be explained by the perceptual processing itself (Rahnev et al., 2011).

One potential limitation of the current method to estimate the cortical depth dependent activity is that the gradient-echo imaging sequence employed here is known to have a better SNR towards the surface of the cortex (De Martino et al., 2018). This SNR difference may mean that the study was underpowered to elucidate potential additional contributions of V1 at the deep cortical depth, although a previous study with similar 7T fMRI parameters has successfully elucidated significant effects in V1 even at its deep cortical depths (Kok et al., 2016). The current results certainly suggest the involvement of V1 in the anticipatory processing of threat, but its layer-specificity may be further investigated in future studies with fMRI sequences rather than gradient-echo.

The notion that the pulvinar plays a role in top-down modulation of fearful face processing is not mutually exclusive with the traditional view emphasizing the role of the pulvinar in the subcortical route ("low road") to process the presented threat signals in an input-driven manner. The subcortical route is thought to bypass the cortex to rapidly relay the retinal input to the amygdala via the superior colliculus and pulvinar (LeDoux, 1996; de Gelder, 2006). Such coarse input-driven processing has been speculated to result in erroneous, false perception of threat signals (LeDoux, 1996). Although the function of such a subcortical route may be relatively degraded in humans compared with other species such as rodents (Pessoa and Adolphs, 2010; Bridge et al., 2016), its contribution has been demonstrated in humans (Tamietto and de Gelder, 2010; Tamietto et al., 2012; Mendez-Bertolo et al., 2016).

Overall, the current results suggest that anticipatory processing of threat signals is distinct from the input-driven processing of threat signals, with the former uniquely involving the engagement of the pulvinar and V1 which constitute the earliest stage of visual processing hierarchy. Such anticipatory processing of threat may contribute to the perception of threat-relevant images in the absence of corresponding sensory inputs as in clinical cases of flashbacks reported in PTSD.

Declaration of interest

We declare no conflict of interest.

eNeuro Accepted Manuscript

eNeuro Accepted Manuscript

References

Andersson JL, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20:870-888.

Bishop SJ (2007) Neurocognitive mechanisms of anxiety: an integrative account. Trends Cogn Sci 11:307-316.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207-4221.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433-436.

Bridge H, Leopold DA, Bourne JA (2016) Adaptive Pulvinar Circuitry Supports Visual Cognition. Trends Cogn Sci 20:146-157.

Calvo MG, Lundqvist D (2008) Facial expressions of emotion (KDEF): identification under different display-duration conditions. Behav Res Methods 40:109-115.

Chakravarty MM, Bertrand G, Hodge CP, Sadikot AF, Collins DL (2006) The creation of a brain atlas for image guided neurosurgery using serial histological data. Neuroimage 30:359-376.

Cicmil N, Krug K (2015) Playing the electric light orchestra--how electrical stimulation of visual cortex elucidates the neural basis of perception. Philos Trans R Soc Lond B Biol Sci 370:20140206.

Cohen BH (2004) Explaining Psychological Statistics. New York: John Wiley & Sons.

Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences. Hillsdale: Lawrence Erlbaum and Associates.

de Gelder B (2006) Towards the neurobiology of emotional body language. Nat Rev Neurosci 7:242-249.

De Martino F, Yacoub E, Kemper V, Moerel M, Uludag K, De Weerd P, Ugurbil K, Goebel R, Formisano E (2018) The impact of ultra-high field MRI on cognitive and computational neuroimaging. Neuroimage 168:366-382.

de Sousa AA, Sherwood CC, Schleicher A, Amunts K, MacLeod CE, Hof PR, Zilles K (2010) Comparative cytoarchitectural analyses of striate and extrastriate areas in hominoids. Cereb Cortex 20:966-981.

Dominguez-Vargas AU, Schneider L, Wilke M, Kagan I (2017) Electrical Microstimulation of the Pulvinar Biases Saccade Choices and Reaction Times in a Time-Dependent Manner. J Neurosci 37:2234-2257.

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815-836.

Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9:416-429.

Grieve KL, Acuna C, Cudeiro J (2000) The primate pulvinar nuclei: vision and action. Trends Neurosci 23:35-39.

Grillon C, O'Connell K, Lieberman L, Alvarez G, Geraci M, Pine DS, Ernst M (2017) Distinct responses to predictable and unpredictable threat in anxiety pathologies: effect of panic attack. Biol Psychiatry Cogn Neurosci Neuroimaging 2:575-581.

Kanai R, Komura Y, Shipp S, Friston K (2015) Cerebral hierarchies: predictive processing, precision and the pulvinar. Philos Trans R Soc Lond B Biol Sci 370.

Kastner S, O'Connor DH, Fukui MM, Fehd HM, Herwig U, Pinsk MA (2004) Functional imaging of the human lateral geniculate nucleus and pulvinar. J Neurophysiol 91:438-448.

Keselman HJ, Keselman JC (1993) Analysis of repeated measurements. In: Applied Analysis of Variance in Behavioral Science (Edwards LK, ed). New York.

Koizumi A, Mobbs D, Lau H (2016) Is fear perception special? Evidence at the level of decision-making and subjective confidence. Soc Cogn Affect Neurosci 11:1772-1782.

664 Kok P, Bains LJ, van Mourik T, Norris DG, de Lange FP (2016) Selective Activation of the Deep
665 Layers of the Human Primary Visual Cortex by Top-Down Feedback. Curr Biol 26:371-
666 376.
667 Lafer-Sousa R, Conway BR (2013) Parallel, multi-stage processing of colors, faces and shapes
668 in macaque inferior temporal cortex. Nat Neurosci 16:1870-1878.
669 Lanzetta JT, Orr SP (1986) Excitatory strength of expressive faces: effects of happy and fear
670 expressions and context on the extinction of a conditioned fear response. J Pers Soc
671 Psychol 50:190-194.
672 LeDoux JE (1996) In: The Emotional Brain New York: Simon and Schuster.
673 Mai JKM, M.; & Paxinos, G. (2015) Atlas of the Human Brain.: Academic Press.
674 Marion R, Li K, Purushothaman G, Jiang Y, Casagrande VA (2013) Morphological and
675 neurochemical comparisons between pulvinar and V1 projections to V2. J Comp Neurol
676 521:813-832.
677 McFadyen J, Mermillod M, Mattingley JB, Halasz V, Garrido MI (2017) A Rapid Subcortical
678 Amygdala Route for Faces Irrespective of Spatial Frequency and Emotion. J Neurosci
679 37:3864-3874.
680 McLaren DG, Ries ML, Xu G, Johnson SC (2012) A generalized form of context-dependent
681 psychophysiological interactions (gPPI): a comparison to standard approaches.
682 Neuroimage 61:1277-1286.
683 Mendez-Bertolo C, Moratti S, Toledano R, Lopez-Sosa F, Martinez-Alvarez R, Mah YH,
684 Vuilleumier P, Gil-Nagel A, Strange BA (2016) A fast pathway for fear in human
685 amygdala. Nat Neurosci 19:1041-1049.
686 Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E (2015)
687 Contextual Feedback to Superficial Layers of V1. Curr Biol 25:2690-2695.
688 Norris DG, Polimeni JR (2019) Laminar (f)MRI: A short history and future prospects.
689 Neuroimage 197:643-649.
690 O'Reilly JX, Woolrich MW, Behrens TE, Smith SM, Johansen-Berg H (2012) Tools of the trade:
691 psychophysiological interactions and functional connectivity. Soc Cogn Affect Neurosci
692 7:604-609.
693 Pajani A, Kok P, Kouider S, de Lange FP (2015) Spontaneous Activity Patterns in Primary
694 Visual Cortex Predispose to Visual Hallucinations. J Neurosci 35:12947-12953.
695 Pessoa L, Ungerleider LG (2004) Neural correlates of change detection and change blindness
696 in a working memory task. Cereb Cortex 14:511-520.
697 Pessoa L, Adolphs R (2010) Emotion processing and the amygdala: from a 'low road' to 'many
698 roads' of evaluating biological significance. Nat Rev Neurosci 11:773-783.
699 Purushothaman G, Marion R, Li K, Casagrande VA (2012) Gating and control of primary visual
700 cortex by pulvinar. Nat Neurosci 15:905-912.
701 Rahnev D, Maniscalco B, Graves T, Huang E, de Lange FP, Lau H (2011) Attention induces
702 conservative subjective biases in visual perception. Nat Neurosci 14:1513-1515.
703 Ress D, Heeger DJ (2003) Neuronal correlates of perception in early visual cortex. Nat Neurosci
704 6:414-420.
705 Roth MM, Dahmen JC, Muir DR, Imhof F, Martini FJ, Hofer SB (2016) Thalamic nuclei convey
706 diverse contextual information to layer 1 of visual cortex. Nat Neurosci 19:299-307.
707 Saalmann YB, Kastner S (2011) Cognitive and perceptual functions of the visual thalamus.
708 Neuron 71:209-223.
709 Schwiedrzik CM, Freiwald WA (2017) High-Level Prediction Signals in a Low-Level Area of the
710 Macaque Face-Processing Hierarchy. Neuron 96:89-97 e84.
711 Shipp S (2003) The functional logic of cortico-pulvinar connections. Philos Trans R Soc Lond B
712 Biol Sci 358:1605-1624.
713 Shipp S (2007) Structure and function of the cerebral cortex. Curr Biol 17:R443-449.
714 Shipp S (2016) Neural Elements for Predictive Coding. Front Psychol 7:1792.

eNeuro Accepted Manuscript

715 Sprenger T, Seifert CL, Valet M, Andreou AP, Foerschler A, Zimmer C, Collins DL, Goadsby PJ,
716 Tolle TR, Chakravarty MM (2012) Assessing the risk of central post-stroke pain of
717 thalamic origin by lesion mapping. Brain 135:2536-2545.
718 Tamaki M, Bang JW, Watanabe T, Sasaki Y (2016) Night Watch in One Brain Hemisphere
719 during Sleep Associated with the First-Night Effect in Humans. Curr Biol 26:1190-1194.
720 Tamietto M, de Gelder B (2010) Neural bases of the non-conscious perception of emotional
721 signals. Nat Rev Neurosci 11:697-709.
722 Tamietto M, Pullens P, de Gelder B, Weiskrantz L, Goebel R (2012) Subcortical connections to
723 human amygdala and changes following destruction of the visual cortex. Curr Biol
724 22:1449-1455.
725 Torrisi S, Gorka AX, Gonzalez-Castillo J, O'Connell K, Balderston N, Grillon C, Ernst M (2018)
726 Extended amygdala connectivity changes during sustained shock anticipation. Transl
727 Psychiatry 8:33.
728 Tottenham N, Tanaka JW, Leon AC, McCarry T, Nurse M, Hare TA, Marcus DJ, Westerlund A,
729 Casey BJ, Nelson C (2009) The NimStim set of facial expressions: judgments from
730 untrained research participants. Psychiatry Res 168:242-249.
731 Van de Moortele PF, Auerbach EJ, Olman C, Yacoub E, Ugurbil K, Moeller S (2009) T1
732 weighted brain images at 7 Tesla unbiased for Proton Density, T2* contrast and RF coil
733 receive B1 sensitivity with simultaneous vessel visualization. Neuroimage 46:432-446.
734 Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level
735 image properties: the SHINE toolbox. Behav Res Methods 42:671-684.
736 Witthoft N, Nguyen ML, Golarai G, LaRocque KF, Liberman A, Smith ME, Grill-Spector K (2014)
737 Where is human V4? Predicting the location of hV4 and VO1 from cortical folding. Cereb
738 Cortex 24:2401-2408.
739
740

eNeuro Accepted Manuscript

741 **Supporting data**

742

743 **Figure 1-1.** Illustration of the differential contribution of anticipation for the HIT and False Alarm (FA)
744 trials. Given that there was no trial-wise cue to forecast the upcoming face stimulus (i.e., fearful or
745 neutral), it is expected that anticipation for the fearful face target was similarly fluctuating across all the
746 trials regardless of the actual type of face stimulus. That is, regardless of the presented face stimulus
747 (fearful or neutral), there are likely to be similar proportions of trials with relatively higher level of
748 anticipation (see inner red circle in the center). When comparing the HIT and FA trials, however,
749 heightened anticipation is likely to contribute to a larger proportion of the FA trials relative to the HIT trials
750 where sensory inputs of presented fearful face also contributed to the percept. This can be inferred when
751 simply considering the proportion of HIT trials that are mainly driven by sensory inputs or alternatively by
752 anticipation. That is, while the proportion of HIT trials (≒ 55%) that is the same as the total FA rate (≒
753 25%) may be attributed to amplified anticipatory processing, the remaining HIT trials (≒ 30% out of 55%)
754 is likely to reflect sensory inputs for the target faces instead. Thus, FA trials are more likely to reflect top-
755 down related processing than are HIT trials. Here, trials are binary divided into higher and lower
756 anticipation trials for simplicity, and the actual proportions may vary across participants.

757
758

759

760 **Figure 2-1.** Pulvinar and V1 activity. **A**. Deconvolved time course for the pulvinar for HIT and false alarm
761 (FA) trials during the detection tasks with fearful and happy face targets (left and right panels,
762 respectively). Error bar indicates standard error of mean. **B**. Deconvolved time course of V1 for the HIT
763 and false alarm (FA) trials during the fearful face detection task and happy face detection task (left and
764 right panels, respectively). Error bar indicates standard error of mean. Related to **Figs 2 and 3**.

765

766

767

eNeuro Accepted Manuscript

768
769

770 **Figure 2-2**. Activity of the pulvinar on FA and HIT trials in each of its subregions. **A.** The subregions
771 composing the lateral, inferior, and medial portions of the pulvinar were defined based on a histological
772 atlas (Chakravarty MM et al. 2006). The atlas was imported to BrainVoyager, and the entire pulvinar
773 including all subregions was manually aligned to the pulvinar in each participant's Native space. We
774 included the subregions that compose the lateral, inferior, and medial portions of pulvinar which are
775 widely implicated in visual processing (Pessoa and Adolphs 2010; Bridge et al. 2016), namely the nucleus
776 pulvinaris oromedialis (lateral), nucleus pulvinaris orolateralis (lateral), pulvinar laterale (lateral), nucleus
777 pulvinaris intergeniculatus (inferior), nucleus pulvinaris (medial), and pulvinar mediale (medial). The entire
778 voxels within each subregion were used to estimate the activity level. **B**. Similarly to the results in the
779 Main text (**Fig. 2**) with the functionally defined pulvinar ROIs, there was significantly larger activity on the
780 FA trials of a fearful face relative to the FA trials of a happy face in two subregions, namely the nucleus
781 pulvinaris oromedialis (lateral) and nucleus pulvinaris (medial). However, there was no significant
782 interaction between emotion (fearful/happy) and condition (HIT/FA) in these two subregions (nucleus
783 pulvinaris oromedialis: $F(1,10) = 4.496$, $p = 0.060$; nucleus pulvinaris: $F(1,10) = 2.327$, $p = 0.158$). The
784 results suggested that one lateral and one medial subregion showed significantly larger activity on FA
785 trials of a fearful face relative to FA trials of a happy face. Given that the same neutral faces were
786 presented on these trials, the difference in activity is likely to be due to the perceived emotion driven by
787 anticipation of a fearful versus happy face target. These results hint at the possibility that task-driven
788 anticipation of threat signals may be coded in the medial pulvinar through its interaction with the prefrontal
789 areas where task-set is generally represented (Lau and Passingham, 2007), while the lateral pulvinar
790 may directly modulate V1 based on anticipation. However, these results should be treated as indicative
791 as the results within these subregions were limited in that there was no significant interaction between
792 conditions (HIT/FA) and emotions (fearful/happy) and that the relatively lower SNR in the inner brain area
793 like the pulvinar may have interfered with the differentiation of subregion activities (see *Estimation of*
794 *tSNRs*). Box plot shows upper (75%) and lower (25%) quartiles with median (red line) and mean (red dot),
795 with whisker showing maximum and minimum value. An outlier (outside of ± 2.7 standard deviations

eNeuro Accepted Manuscript

within a distribution for a given condition) is shown with a red cross. *$p < 0.05$ in planned post-hoc tests. r: right, l: left, a: anterior, p: posterior. Related to **Fig. 2**.

eNeuro Accepted Manuscript

804

805 **Figure 3-1.** V1 activity in each participant at superficial (top row) and deep cortical depths (bottom row).
806 For each subject, deconvolved time course of V1 during the HIT and false alarm (FA) trials during the
807 fearful face detection task and happy face detection task (left and right panels, respectively) are shown.
808 Subjects who showed numerically larger activity on FA trials of a fearful face relative to FA trials of a
809 happy face, which is in line with the pattern of group level result shown in **Fig. 3C**, are highlighted with red
810 fonts (8 out of 11 participants). Related to **Fig. 3**.

811

812

813

814

815

816

817 **Figure 3-2.** The results of gPPI analyses examining the connectivity between the pulvinar and V1**.** gPPI
818 analyses were separately conducted, with the V1 superficial or deep cortical depth voxels as the seed
819 ROI (see G*eneralized form of context-dependent psychophysiological interaction analysis (gPPI)*). Mean
820 t-values for the parameter estimates in gPPI for each trial type and seed ROI are shown. Only during FA
821 trials with fearful faces, a significant modulation of connectivity was present between the pulvinar and V1
822 superficial layers (t(10) = 3.981, p = 0.0026, one-sample t-test against 0, significant after Bonferroni
823 correction). We note that these results are only indicative, as there was no significant interaction between
824 conditions (HIT/FA), facial emotion (fearful/happy), and V1 cortical depth (superficial/deep) (F(1,10)=.198,
825 p = .666). There was no evidence that the connectivity between the pulvinar and V1 was modulated
826 during other conditions in either V1 cortical depths. With the V1 superficial cortical depth as a seed ROI,
827 fearful HIT: t(10) = 0.403, p = 0.695, CI [-0.683 0.985]; fearful FA: t(10) = 3.981, p = 0.0026, CI [0.319
828 1.128]; happy HIT: t(10) = 1.869, p = 0.091, CI [-0.105 1.120]; happy FA: t(10) = 1.4364, p = 0.181, CI [-
829 0.271 1.253]. With the V1 deep cortical depth as a seed ROI, fearful HIT: t(10) =0.398 , p = 0.699, CI [-
830 0.641 0.92]; fearful FA: t(10) = 2.223, p = 0.05, CI [-0.001 1.199]; happy HIT: t(10) = 1.594, p = 0.142, CI
831 [-0.15 0.01]; happy FA: t(10) = 1.152, p = 0.276, CI [-0.459 1.441]. We note that the null results do not

eNeuro Accepted Manuscript

832 necessarily indicate the absence of effects, because a PPI analysis is generally low in power (O'Reilly et
833 al., 2012) and tSNR in the pulvinar was relatively low in the current study. Future studies may further
834 examine the potential changes in the functional connectivity with the pulvinar across the V1 cortical
835 depths and experimental conditions. pulv: pulvinar, a: anterior, p: posterior.
836
837
838
839
840
841
842
843
844
845

eNeuro Accepted Manuscript

Figure 3-3. Prior to the onsets of target faces, there was already enhanced activity in the pulvinar preceding the false alarm (FA) trial of a fearful face compared with a happy face (left panel). To estimate activity prior to the face onsets, the z-normalised signal change (%) in the preprocessed raw time course (see *fMRI processing*) was averaged between the 2 time points immediately before the face onsets (i.e., -5 and -2.5 s) relative to the preceding baseline (averaged between the 2 earlier time points, i.e., -10 and -7.5 s). A repeated measures ANOVA with two factors of percept type (pre-HIT/pre-FA) and emotion (fearful/happy) revealed a significant interaction ($F(1,10) = 10.094$, $p = 0.010$). Post-hoc analyses revealed that pulvinar activity was significantly greater on FA trials of fearful faces compared with happy faces in the pre-onset period ($t(10) = 3.392$, $p = 0.007$, $d = 1.1$). During the same pre-onset period, while pulvinar activity was significantly larger on HIT than on FA trials of happy faces ($t(10) = 2.386$, $p = 0.038$, $d = 0.7$), there was a non-significant opposite trend such that activity was relatively greater on FA than on HIT trials of fearful faces ($t(10) = -2.147$, $p = 0.057$, n.s.). Meanwhile, in the same pre-onset period, there was not yet any differential activity in the V1 superficial cortical depth (right panel), with no significant main effects or interaction between percept type and emotion ($ps > 0.50$, n.s.). These results suggest that anticipatory activity in the pulvinar modulated the response of the V1 superficial cortical depth triggered by the subsequent onset of a face target shown in **Fig. 3D**. Box plot shows upper (75%) and lower (25%) quartiles with median (red line) and mean (red dot), with whisker showing maximum and minimum value. $^{**}p < 0.01$, $^{*}p < 0.05$, $^{+}p < 0.10$. Related to **Figs 2 and 3**.

Figure 3-4. The results of a control analysis showing no differential activity in the pulvinar and V1 on MISS versus Correct rejection (CR) trials. We examined whether similar results on FA trials of fearful faces (**Figs 3B and 4D**) may be present on MISS trials, where a fearful face was presented but was not detected. While the results for FA trials of fearful faces may reflect enhanced top-down processing, other non-mutually exclusive possibilities are worth considering. Specifically, the results for

eNeuro Accepted Manuscript

FA trials may reflect a mere mismatch between sensory input and reported percept. Although this possibility is unlikely considering the fact that such results did not generalize to FA trials of happy faces, there remains a possibility that such input-to-percept mismatch may evoke certain neural activity only in anticipation of threat cues. We therefore examined whether the patterns of results we obtained from the contrast between FA and HIT trials may be also obtained from the contrast between MISS and correct rejection (CR) trials, where neutral faces were perceived with or without mismatched sensory input, respectively. **A.** Unlike the contrast of Pulvinar activity between the HIT and false alarm (FA) trials shown in **Fig. 2**, there was no difference in pulvinar activity between the MISS and CR trials regardless of the facial emotion (a non-significant interaction between percept type (MISS/CR) and emotion: $F(1,10) = 0.013$, $p = 0.912$). Although we saw a trend for pulvinar activity to be larger for the fearful face detection task than for the happy face detection task, there was no significant main effect of emotion ($F(1,10) = 2.365$, $p = 0.155$). **B.** Similarly, unlike the contrast of V1 activity between the HIT and FA trials shown in **Fig. 3**, there was no differential activity in V1 between the MISS and CR trials, regardless of emotion and cortical depth (a non-significant second-order interaction: $F(1,10) = 1.251$, $p = 0.290$). While there was generally greater activity on MISS trials relative to CR trials that was not specific to any facial emotion and potentially reflected sensory inputs of salient emotional faces albeit undetected, there was no significant main effect of percept type ($F(1,10) = 2.573$, $p = 0.140$). Box plot shows upper (75%) and lower (25%) quartiles with median (red line) and mean (red dot), with whisker showing maximum and minimum value. An outlier (outside of ± 2.7 standard deviations within a distribution for a given condition) is shown with a red cross. Related to **Figs 2B and 3D**.

eNeuro Accepted Manuscript