Is fear perception special? Evidence at the level of decision-making and subjective confidence

Ai Koizumi,1,2,3,4 Dean Mobbs,1,5 and Hakwan Lau6

1Department of Psychology, Columbia University 406 Schermerhorn Hall, 1190 Amsterdam Ave MC 5501, New York, NY 10027, USA, 2US-Japan Brain Research Cooperation Program, National Institute for Physiological Sciences, Japan, 3Japan Society for the Promotion of Science (JSPS), Tokyo, Japan, 4Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology, 1-4 Yamadaoka, Suita City, Osaka 565-0871, Japan, 5California Institute of Technology, Humanities and Social Sciences, Baxter Hall 1200 E. California Blvd. Pasadena, CA 91125, USA, and 6Department of Psychology & Brian Research Institute, UCLA 1285 Franz Hall, Bo 951563, Los Angeles, CA 90095-1563, USA

Correspondence should be addressed to Ai Koizumi, Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology, 1-4 Yamadaoka, Suita City, Osaka 565-0871, Japan. E-mail: bellkoizumi@gmail.com

Abstract

Fearful faces are believed to be prioritized in visual perception. However, it is unclear whether the processing of low-level facial features alone can facilitate such prioritization or whether higher-level mechanisms also contribute. We examined potential biases for fearful face perception at the levels of perceptual decision-making and perceptual confidence. We controlled for lower-level visual processing capacity by titrating luminance contrasts of backward masks, and the emotional intensity of fearful, angry and happy faces. Under these conditions, participants showed liberal biases in perceiving a fearful face, in both detection and discrimination tasks. This effect was stronger among individuals with reduced density in dorsolateral prefrontal cortex, a region linked to perceptual decision-making. Moreover, participants reported higher confidence when they accurately perceived a fearful face, suggesting that fearful faces may have privileged access to consciousness. Together, the results suggest that mechanisms in the prefrontal cortex contribute to making fearful face perception special.

Key words: fearful face perception; perceptual decision-making; metacognition; dorsolateral prefrontal cortex (DLFPC); voxel-based morphometry (VBM)

Introduction

Many studies suggest that fearful face perception is special, in that a fearful face is easier or quicker to be perceived than faces with neutral or other emotional expressions. For instance, some studies have shown that fearful faces were more likely to be perceived when attention was directed away (e.g. attentional blink; Milders et al., 2006; Stein et al., 2009). Other studies have shown that fearful faces are more resistant to suppression by visual noise or dichoptic suppression (Yang et al., 2007; Amting et al., 2010; Stein et al., 2010, 2014; Stienen and de Gelder, 2011). Yet, it is unclear if such prioritized perception of fearful faces is merely due to lower-level visual processing. Although previous studies carefully controlled for low-level physical features such as luminance contrast and spatial frequency, bottom-up processing of local features such as the increased eye white exposure may be sufficient to explain why fearful faces are easier to be seen. For instance, Yang et al. (2007) showed that fearful faces resisted continuous flash suppression better than other faces. But
importantly, the effect was preserved even when the faces were inverted. These results seem to be compatible with the interpretation that it was local features rather than Gestalt facial processing that account for the prioritized perception of fearful faces.

On the other hand, it is also clear that high-level mechanisms contribute to visual perception in general. For example, it has been shown that mechanisms in the prefrontal cortex contribute to our perception by forming perceptual decisions, that is to statistically evaluate the most likely identity of the perceived object given the accumulating sensory evidence encoded in lower sensory areas (Gold and Shadlen, 2007). This highlights the importance of examining high-level processes in an attempt to thoroughly understand the mechanism underlying our visual perception. Yet, while the low-level mechanisms have been the focus in previous studies (Yang et al., 2007), the contribution of high-level mechanisms in the fearful face perception has been rarely investigated. We speculated that these high-level mechanisms may also contribute to prioritize fearful face perception, by biasing perceptual decisions in favor of fearful faces.

Another high-level mechanism involved in visual perception is metacognition, with which we monitor our ongoing perceptual decisions and derive a level of subjective confidence when we are likely to be correct. It has been shown that fearful faces are more likely to reach conscious perception (Yang et al., 2007; Amting et al., 2010; Stein et al., 2010, 2014; Stienen and de Gelder, 2011), which in turn has been linked to metacognition (Lau and Rosenthal, 2011; Ko and Lau, 2012). Thus, we also speculated that fearful faces may have a prioritized access to the metacognitive system, such that fearful faces may be perceived with higher confidence.

The possibility that high-level mechanisms contribute to prioritize fearful face perception is not incompatible with low-level accounts. Given that fearful faces could forecast approaching dangers in the environment, there may be a survival advantage to prioritize perception of fearful faces at both the higher and lower levels.

Therefore, in this study, we examined potential biases for fearful face perception at the levels of perceptual decision-making and metacognition. To selectively examine biases due to high-level mechanisms, we carefully controlled for lower-visual processing sensitivity by titrating the luminance contrasts of backward masks, as well as the emotional intensity expressed in fearful, angry and happy faces. We also used voxel-based morphometry (VBM) to examine the involvement of dorsolateral prefrontal cortex (DLPFC), which plays an important role in both perceptual decision-making (Kim and Shadlen, 1999; Leon and Shadlen, 1999; Heekeren et al., 2006; PhiliaStides et al., 2011) and metacognition (Lau and Passingham, 2006; Lau and Rosenthal, 2011; Fleming et al., 2012).

### Methods

#### Participants

Forty-one students (17 males, mean age 23.2) from Columbia University participated in this study. All participants were recruited to complete two sessions conducted on two separate days, once with the male image set and once with the female image set (order counterbalanced; see Stimuli for details). Yet, only 24 participants completed both sessions. For the remaining 17 participants, 13 were tested only with the female face image set and 4 were tested only with the male face image set. Data collection was stopped after the number of participants who completed both sessions met the sample size which was predetermined based on the previously reported similar psychophysics experiments (Koizumi et al., 2015). In 24 out of the 41 participants (14 completed both image sets, 2 completed only the male image set and 8 completed only the female image set), we had access to structural MRI images of their brains (10 males, mean age 23.7) from another experiment not mentioned here, thus their data were further used in VBM analysis. All participants received $10–15 per hr of participation. Participants completed a written informed consent form prior to participation. The study was approved by the Columbia University’s Committee for the Protection of Human Subjects.

#### Stimuli

We used the NimStim stimulus set (Tottenham et al., 2009) to select the images of fearful, angry, happy and neutral faces, posed by one male and one female models (model number 34M and 07F) (to demonstrate in Figure 1, another model (03F) with publication permission was used). Only the images with open mouth were selected so that participants could not perform the tasks solely based on whether the mouth was open in that image. The images were morphed from neutral to fearful, angry and happy expressions of the same model using the NorrKross MorphX software version 2.11. Each morph sequence (e.g. neutral to fear) had 99 intervals, producing a continuum of 100 images with the gradual increment of emotional intensity. The morphed images were converted into gray scale, and their luminance contrast, spatial frequency spectrum and overall luminosity were equated across the entire image set with the SHINE toolbox (Willebockel et al., 2010) implemented in Matlab. The image contrast was 30% of the original images. The size of the image was 8.4 × 6.5 in visual angle with a viewing distance of 75 cm.

The face image was followed by a mask (see Experimental Design, Figure 1) so as to manipulate the task difficulty but the mask presentation was not meant to completely abolish awareness of the face. The mask images were created by modifying the to-be-masked face images from that trial in a manner that the phase of face stimuli was scrambled using Matlab while maintaining other physical properties (e.g. luminosity and spatial frequency spectrum).

**Experimental design**

**Detection calibration.** Participants first went through Detection Calibration, in which they performed fearful, angry and happy face detection tasks, in orders counterbalanced across participants. The aim of calibration was to identify a threshold-level intensity for each emotion so that the participants would perform equally well (i.e. same sensitivity) in detecting faces with three different emotions. This allowed us to access detection biases while making sure that such biases were not due to differences in the emotional intensity between the faces. During calibration, we titrated the emotional intensity of the target face (see Stimuli earlier) with the Quest threshold estimation procedure (Watson and Pelli, 1983), so that the participants were expected to perform at an accuracy of about 75% in a 2-interval forced-choice (2IFC) detection task (details later). Each task had 16 practice trials and 96 calibration trials. The calibration trials consist of two randomly interleaved Quest sequences of 48 trials each. Each Quest sequence independently estimated the threshold emotional intensity, and the average of the two estimates was used in the following Discrimination Calibration and experimental tasks. The means of morph intensity for fear, angry and happy face were $M = 48.63 ± s.d. 27.21$, $M = 51.38 ± 27.29$, $M = 40.23 ± 19.50$, respectively. A one-way
analysis of variance (ANOVA) revealed a significant main effect of Emotion \[F(2, 39) = 6.88, P = 0.003, partial \eta^2 = 0.26\], which was due to that the calibrated morphed intensity was higher for an anger face \((P = 0.002, \text{Bonferroni corrected})\) and tended to be higher for a fearful face \((P = 0.06)\), relative to a happy face.

On each trial, one neutral face as well as one target emotional face (fearful, happy or angry, in different blocks) were presented consecutively in random order. A trial started with a face (33 ms) followed by a mask (99 ms). After a blank (750 ms), a second face (33 ms) was followed by another mask (99 ms). The mask contrast was matched with the contrast of face image and was kept constant throughout Detection Calibration. To avoid the overlap of locations between the two faces, the location of the first face was jittered from the center of screen to a random direction by a Euclidean distance of 1.2 degree, and the location of the second face was jittered by the same distance to a different direction. Immediately after the second mask, the key-assignment appeared in the lower screen. Participants used 1- or 2-key to report the perceived sequence of faces, i.e. whether the emotional face came first, or the neutral face came first. The key assignment was counterbalanced across the participants. When a response was registered or a 5 s period elapsed from the offset of the second mask, a fixation cross (0.35°) was presented for 250 ms, followed by the next trial.

**Discrimination calibration.** With the emotional intensity levels determined in Detection Calibration, participants went through Discrimination Calibration in which they performed three tasks to discriminate fearful/angry, fearful/happy and happy/angry emotion pairs, in orders counterbalanced across participants. The aim of this additional calibration was to identify a threshold contrast for the backward mask for each emotion pair, so that the participants would discriminate each pair with similar accuracy. The reason is that even when each emotional face is similarly detectable, discrimination between each face pair may still show different sensitivities. The luminance contrast of the backward mask was titrated with the Quest threshold estimation procedure (Watson and Pelli, 1983) to achieve an accuracy of 75% in a 2IFC task. The procedure was similar to Detection Calibration. In each calibration task, a total of 96 trials were divided into two randomly interleaved Quest sequences of 48 trials. Each sequence independently estimated the threshold mask contrast, and the average of the two estimates was used in the experimental Discrimination Tasks.

The means of calibrated Michelson contrast for the backward mask in the discrimination tasks with fearful/angry, fearful/happy and happy/angry pairs were \(M = 34.06 \pm \text{s.d. 13.60,} \ M = 26.96 \pm 13.82, \ M = 34.62 \pm 13.78, \) respectively. A one-way ANOVA revealed a significant main effect of Emotion \([F(2, 39) = 5.65, P = 0.007, partial \eta^2 = 0.23]\), which was due to that the mask contrast for the discrimination task was significantly lower for the fearful/happy pair compared with the fearful/angry pair as well as the happy/angry pair \((P = 0.033, P = 0.019, \text{respectively})\).

**Detection tasks.** With the estimated threshold intensity for each emotion, participants went through three Detection Tasks (fearful/neutral, angry/neutral and happy/neutral) and three Discrimination Tasks (fearful/angry, fearful/happy and happy/angry), in orders counterbalanced across participants. The mask with the same contrast as the face images were used in the Detection Tasks, and the mask contrasts estimated in Discrimination Calibration were used in the Discrimination Tasks. In Detection Tasks (Figure 1A), participants were presented with either a neutral face or an emotional face expressing fearful, angry or happy emotions. Each trial started with a face (33 ms) followed by a mask (99 ms). Immediately after the mask, a key assignment was shown in the lower portion of the screen. Participants used 1- or 2-key to indicate which of the two face images (e.g. fearful or neutral) was perceived. The key assignment was counterbalanced across participants. After 200 ms from the response, a key assignment for confidence rating was indicated in the lower portion of the screen. Participants rated how confident they were that their response was accurate. They used the 1- to 4-keys on a keyboard to indicate from lowest to highest confidence. When confidence was rated or 5 s elapsed from the mask offset, a fixation cross (0.35°) was presented for 250 ms, followed by the next trial. The location was matched for the face and mask images within a trial but was jittered by a distance of 1.2 degree in a random direction across trials.

**Discrimination tasks.** The procedure for the Discrimination Tasks (Figure 1B) was similar to those of the Detection Tasks. On each trial, participants were presented with one of the face images from the emotion pair in that task (e.g. fearful/happy). They indicated their perceptual responses (e.g. fearful) and confidence levels in the same manner as in the Detection Tasks.

**Analyses.**

**Behavioral analyses.** To assess perceptual sensitivity, \(d'\) was calculated for the Detection Tasks and Discrimination Tasks with
standard signal detection theory (SDT) methods (Macmillan and Creelman, 2005). We used d’ values, rather than the more commonly used d’, because d’ is adjusted for the unequal variances of the internal response distributions for the target (e.g. fearful face) and noise (e.g. neutral face), which is often observed in detection tasks (Macmillan and Creelman, 2005). Criterion measure, c, was also calculated for Detection Tasks and Discrimination Tasks with SDT methods (Macmillan and Creelman, 2005). In the subsequent analyses, the performance with the female and male image sets were combined for the participants who completed the experiment with both image sets, as they were generally similar to each other (Supplementary Tables S1–S3) unless otherwise stated in Results.

To examine metacognitive performance in the Discrimination Tasks, we calculated meta-d’ for each response type (e.g. response of perceiving a fearful face), which quantifies how well confidence ratings track perceptual accuracy over trials (Maniscalco and Lau, 2012, 2014) with a code implemented in Matlab (http://www.columbia.edu/~bsm2105/Type2detfit_rs_meta_d_MLE.m). Meta-d’ was not calculated for the Detection Tasks, given that there remain some methodological issues to be solved in such calculation (Maniscalco and Lau, 2014), namely that the unequal variance assumption mentioned earlier is unlikely to be met for detection tasks.

### VBM analyses

Structural brain images were preprocessed with SPM8 (http://www.fil.ion.ucl.ac.uk/spm) in a similar manner as previous studies (Fleming et al., 2010; McCurdy et al., 2013): the T1 scan images were segmented into gray matter, white matter and cerebrospinal fluid. To enhance accuracy of intersubject alignment, the gray matter images were aligned and warped to a template which was iteratively modified with the VBM diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL) algorithm (Ashburner, 2007). This DARTEL template was registered to Montreal Neurological Institute stereotactic space, and the gray matter images were modulated while their tissue volumes were preserved. A Gaussian kernel of 8 mm full width at half-maximum was used to smooth the images.

Given our a priori hypotheses, we selected regions of interest (ROIs) in DLPFC and Amygdala (Figure 4A). DLPFC was selected based on previous literature showing the brain sites important for perceptual decision-making (Heekeren et al., 2006; Philatides et al., 2011). Amygdala was selected based on the literature showing its primary role in fearful face perception (Morris et al., 1996; Adolphs, 2008; Fusar-Poli et al., 2009). The ROI for Amygdala was defined using wfu PickAtlas Tool (http://fmr.i.wfubmc.edu/software/PickAtlas). For the localization of DLPFC, the masks covering Brodmann’s areas (BA) 9 and 46 were created separately for the left and right hemispheres, using wfu PickAtlas Tool. Gray Matter density within each ROI was calculated and extracted using SPM8 and Matlab (7.10.0). The calculated values were adjusted for the effects of total brain volume, gender and age and were sent to the subsequent analyses.

### Results

#### Detection tasks

First we analyzed perceptual sensitivity, d’ (Figure 2A), in three Detection Tasks with fearful, angry and happy faces. A one-way ANOVA with Emotion (fearful, angry and happy) as a within-subject-factor showed no significant main effect [F(2, 39) = 2.32, P = 0.11, partial η² = 0.11], indicating that detection d’ did not differ across the tasks. This means that perceptual sensitivity was successfully equated in Detection Calibration.

For detection criteria (Figure 2B), a one-way ANOVA showed a significant main effect of Emotion [F(2, 39) = 7.05, P = 0.002, partial η² = 0.27]: criteria in the fearful face Detection Task was more liberal (i.e. smaller value) than in the angry face Detection Task (P = 0.04) and happy face Detection Task (P = 0.002). This liberal bias in fearful face detection was also apparent in the analyses of hit rate (HR) and false alarm rate (FAR) (Figure 2C and D, respectively). For HR, a one-way ANOVA revealed a significant main effect of Emotion [F(2, 39) = 4.73, P = 0.01, partial η² = 0.20], which was due to that HR was higher in detecting a fearful face than a happy face (P = 0.02).

Likewise, for FAR, there was a significant main effect of Emotion [F(2, 39) = 4.68, P = 0.02, partial η² = 0.19], which was due to that FAR in detecting a fearful face was higher than detecting an angry face (P = 0.01).

Taken together, these results show that participants were more likely to report perceiving a ‘fearful’ face in the context of fearful face detection, regardless of whether a fearful face was physically presented.

Given one earlier study (Rahnev et al., 2011), one could suggest a somewhat trivializing account for the currently observed liberal bias in fearful face detection: Rahnev et al. (2011) showed that liberal criteria can be due to a large trial-by-trial variance of the internal perceptual response. If we assume the detection criterion is constant across task conditions, when the variance of internal response is large, this could mean that the response may frequently cross the detection criterion, leading to many ‘yes’ responses. Thus the observed liberal criterion for fearful face detection may be because fearful faces were associated with higher variance than the internal perceptual response.

However, analyses on the internal response variability did not support this account (Rahnev et al., 2011). Specifically, we examined the degree to which the variance of target distribution (emotional faces) was different from the variance of noise distribution (neutral faces) in each Detection Task. The ratio of the two variances is reflected by slope of the standardized receiver operating characteristics (z-ROC) curves (Macmillan and Creelman, 2005). Thus, a slope value smaller than 1 indicates a larger variance for the signal than for the noise distribution. A one-way ANOVA on the estimated slope revealed a marginally significant main effect of Emotion [F(2, 39) = 2.70, P = 0.08]. This effect was due to that the slope relatively larger (i.e. closer to 1, where slope = 1 indicates equal variance), rather than smaller, in the fearful face Detection Task (M = 0.93 ± 0.29) compared with the Detection Tasks with happy faces (M = 0.83 ± 0.34) and angry faces (M = 0.89 ± 0.24), although the differences were not statistically significant (P = 0.75 and 0.49, respectively). These results suggest that the internal response for a fearful face had a relatively smaller variance, rather than a larger variance, compared with the responses for happy and angry faces. These results thus suggest that a simple account based on differences in variances of internal response between the face emotions is unlikely.

Although the slope (i.e. the degree of unequal variance) differed across the Detection Tasks, the slope was significantly smaller than 1 in all three tasks (P < 0.05), where slope = 1 indicates the equality of variance between the signal and noise. This result means that the variance of internal signal was larger for the signal (e.g. fearful face) than for the noise (i.e. neutral face) in all three tasks, which is a typically observed trend in a detection task (Macmillan and Creelman, 2005). That all of
current Detection Tasks indeed had an unequal variance between the signal and noise highlights the need to adopt \( d' \) rather than \( d'' \), as a measure of sensitivity in our data set (Macmillan and Creelman, 2005).

To further ensure that the liberal bias in fearful face perception was not due to differences in variance for the internal responses for the different face emotions, we examined criteria in fearful face perception in the Discrimination Tasks. As we will report later, the variances between the two stimuli categories were generally equal in the Discrimination tasks.

As expected from the purpose of calibration to modulate the perceptual sensitivity, calibrated morphing intensity of emotions showed correlation with perceptual sensitivity (\( d' \)) in the Detection tasks. Higher intensity of a happy face was positively correlated with \( d' \) in the Detection task for a happy face (\( R = 0.57, P = 0.004 \)), and higher intensity of an angry face was positively correlated with \( d' \) in the Detection task for an angry face (\( R = 0.66, P < 0.001 \)). Additionally, \( d' \) in the angry Detection task was positively correlated with the intensity of a fearful face (\( R = 0.49, P = 0.016 \)). Meanwhile, the calibrated morphing intensity did not correlate with the criteria (\( P > 0.16 \)).

Fig. 2. Results from Detection Tasks. (A) Detection \( d' \) was similar across the three Detection Tasks (fearful, angry and happy). (B) Criteria for detecting a fearful face was significantly more liberal (i.e. smaller value) than criteria for detecting an angry or happy face. (C) Hit rate was higher when detecting a fearful than a happy face. (D) FAR was significantly higher when detecting a fearful face than an angry face. (**\( P < 0.01 \), *\( P < 0.05 \)).

Discrimination tasks

For Discrimination \( d' \) (Figure 3A), a one-way ANOVA with Emotion Pair as a within-subject factor showed a significant main effect [\( F(2, 39) = 7.03, P = 0.002, \eta^2 = 0.27 \)]. This was due to that \( d' \) was significantly lower in the happy/angry Discrimination Task than in the fearful/happy Discrimination Task (Bonferroni adjusted, \( P = 0.002 \)). Meanwhile, \( d' \) in the fearful/happy Discrimination task did not significantly differ from the fearful/angry Discrimination tasks (\( P = 0.80 \)). There was no difference in \( d' \) between the happy/angry and fearful/angry Discrimination tasks (\( P = 0.18 \)).

For discrimination criteria (Figure 3B), another one-way ANOVA showed a main effect of Emotion Pair [\( F(2, 39) = 7.03, P = 0.002, \eta^2 = 0.27 \)]. To further examine this main effect, we analyzed whether the mean criteria in each Discrimination Task significantly differed from 0, where criteria = 0 corresponds to no-bias (i.e. equally likely to make two alternative responses in a given task, such as ‘fearful’ and ‘angry’). This analysis revealed that participants showed a significant bias to report perceiving a fearful face (i.e. criteria smaller than 0 in Figure 3B) in both Discrimination Tasks that involved a fearful
face (fearful/angry pair, \( P = 0.007 \); fearful/happy pair, \( P < 0.001 \)). Meanwhile, the analysis showed a marginally significant bias to report perceiving a ‘happy’ face in the happy/angry Discrimination Task (\( P = 0.052 \)). Although the results for discrimination criteria were qualitatively similar between the male and female image sets, it was predominantly observed with the male image set (Supplementary Table S2). To further demonstrate the observed bias toward fearful face perception, HR and FAR (Figure 3C and D, respectively) were averaged across the three Discrimination Tasks. Here, HR for perception of a fearful face was calculated as the proportion of a fearful face response when a fearful face was presented. Similarly, HR for perception of an angry and happy face was each calculated as the proportion of an angry or happy face response when an angry or happy face was presented, respectively. That is,

\[
\text{HR}_{\text{fearful}} = P(\text{respond fearful} \mid \text{fearful face}),
\]

\[
\text{HR}_{\text{angry}} = P(\text{respond angry} \mid \text{angry face}),
\]

\[
\text{HR}_{\text{happy}} = P(\text{respond happy} \mid \text{happy face}).
\]

Likewise, FAR for perception of a fearful, angry and happy face was each calculated as

\[
\text{FAR}_{\text{fearful}} = P(\text{respond fearful} \mid \text{non-fearful face}),
\]

\[
\text{FAR}_{\text{angry}} = P(\text{respond angry} \mid \text{non-angry face}),
\]

\[
\text{FAR}_{\text{happy}} = P(\text{respond happy} \mid \text{non-happy face}).
\]

An one-way ANOVA on HR revealed a significant main effect of Emotion \( F(2, 39) = 14.80, P < 0.001, \) partial \( \eta^2 = 0.43 \). Multiple comparisons revealed that HR_{fearful} was significantly higher than HR_{angry} and HR_{happy} (\( P < 0.01 \)). For FAR, there was also a significant main effect of Emotion \( F(2, 39) = 7.82, P = 0.001, \) partial \( \eta^2 = 0.29 \), which was due to that FAR_{fearful} and FAR_{angry} were significantly higher than FAR_{happy} (\( P = 0.002, P = 0.006, \) respectively). Collectively, these results suggest that participants were likely to report perceiving a fearful face, when a fearful face was actually presented (i.e. hit) as well as when an angry or happy face was presented (i.e. false alarm).

Although d’a was relatively higher for the fearful/happy Discrimination task, this unlikely explains the results in criteria as the liberal bias in perceiving a fearful face was similarly observed in the fearful/happy as well as in the fearful/angry Discrimination tasks.

Unlike in the Detection Tasks, the slope of the ROC curves in each Discrimination Task did not statistically differ from 1, where slope = 1 corresponds to the equality of variance (\( P > 0.40 \)).

The calibrated contrast of a backward mask showed correlation with sensitivity (d’a) in some of the Discrimination task. The contrast of mask for the happy/angry Discrimination task was
negatively correlated with its d’a (R = −0.75, P < 0.001). Similarly, the mask contrast for the fearful/happy task showed a negative correlation with d’a in the fearful/happy Discrimination task as well as in the fearful/angry Discrimination task (R = −0.51, P = 0.012; R = −0.50, P = 0.013). The mask contrast in the fearful/angry Discrimination task was also negatively correlated with its discrimination d’a (R = −0.701, P < 0.001). The mask contrast did not correlate with the Discrimination criteria. The calibrated morphing intensity of facial emotion did not correlate with Discrimination d’a and criteria, except that the morphing intensity of a happy face showed positive correlation with the criteria in the fearful/happy Discrimination (R = 0.55, P = 0.005), indicating that participants with stronger calibrated intensity of a happy face was more liberal to report perceiving a happy face.

**Voxel-based morphometry**

The earlier analyses revealed a liberal bias to report perceiving a fearful face in the Detection Tasks as well as in the Discrimination Tasks. Next, we examined whether such bias would correlate with the brain anatomy related with higher-level cognitive processing. To do so, we analyzed whether liberal criteria in fearful face perception correlates with the individual differences in the gray matter density of DLPFC, which is an important region for perceptual decision-making (Heekeren et al., 2006; Philiaistides et al., 2011). The correlations of perceptual bias with left and right DLPFC (BA46 and 46) were tested separately, given previous literatures suggesting distinctive cognitive functions of DLPFC between the hemispheres (Kaller et al., 2011; Barbey et al., 2013). We also analyzed whether such liberal criteria would correlate with the gray matter density of bilateral Amygdala, which is a critical site for fearful face perception (Morris et al., 1996; Adolphs, 2008; Fusar-Poli et al., 2009). In the subsequent analyses, we controlled for perceptual sensitivity (d’a) so as to specifically examine the neural correlates of criteria, by first removing the effects of sensitivity in regression analyses. Specifically, we regressed out d’a from the criteria in the corresponding task. That is, we regressed out d’a in the fearful face Detection task from the criteria in the corresponding fearful face Detection task. Likewise, we regressed out d’a in each of the Detection tasks with an angry and happy face from the criteria in the corresponding Detection tasks, and we then averaged the residual criterion values between the two Detection tasks.

First, we examined the brain structural correlates of criteria in the Detection Tasks with left DLPFC. As Figure 4B shows, reduced DLPFC (BA46L) density was significantly related with more liberal criteria (i.e. <0) in detecting a fearful face (R = 0.414, P = 0.044). Meanwhile, less dense DLPFC (BA46L) did not correlate with the criteria in detecting a non-fearful (angry or happy) face (R = −0.218, P = 0.306), which was the case even when the criterion results were separately analyzed for an angry face (R = −0.01, P = 0.95) and a happy face (R = −0.35, P = 0.10). The two correlations, one with the fearful face detection criteria and the other with the non-fearful face detection criteria, significantly differed from each other [t(21) = −3.10, P = 0.005], as revealed by a William’s t-test (Williams, 1959; Weaver and Wuensch, 2013). This means that participants with reduced DLPFC (BA46L) density were more likely to report perceiving a fearful face, while they were less likely to report perceiving an angry or happy face. Yet, the correlation between BA46L and liberal criteria in detecting a fearful face should be treated as a tentative result, given it would not survive a correction for multiple comparisons (i.e. a threshold is to be adjusted as 0.025 when compensating for testing with BA46L and BA9L). BA9L did not show significant correlation with neither the criteria in detecting a fearful face nor a non-fearful face (R = 0.24, P = 0.27; R = −0.04, P = 0.85, respectively).

When focusing on right DLPFC, both right BA46R and BA9R showed numerically similar results as BA46L, yet they were not statistically significant.

Meanwhile, the Amygdala density correlated with the fearful face detection criteria in the opposite manner as the DLPFC density (Figure 4C). That is, higher Amygdala density, rather than lower density, was related with more liberal criteria for detecting a fearful face (R = −0.419, P = 0.041). Amygdala density did not correlate with criteria for detecting non-fearful faces (R = −0.082, P = 0.704), which was the case even when the results were separately analyzed for an angry and happy face (R = −0.24, P = 0.26; R = 0.11, P = 0.60, respectively). The two correlations of the Amygdala density with a fearful and non-fearful faces did not significantly differ from each other [t(21) = 1.38, P = 0.181].

The correlations of detection criteria with BA46L and Amygdala density mentioned earlier were qualitatively similar even when d’a was not controlled for: lower BA46L density was related with bias toward detecting a fearful face (R = 0.32, P = 0.13), while it was not the case when detecting non-fearful faces (R = −0.23, P = 0.28). Meanwhile, there was a trend that higher Amygdala density was related with bias toward perceiving a fearful face (R = −0.36, P = 0.09), although there was no such correlation between Amygdala density and bias toward perceiving non-fearful faces (R = −0.07, P = 0.74).

The same correlation analyses for the criteria in Discrimination Tasks (Supplementary Figure S1) showed similar results as the Detection Tasks (Figure 4), although the results were not statistically significant.

The densities of BA46L and Amygdala were not related with the calibrated stimulus properties, as shown by lack of their correlations with the morphing intensity of facial emotions (P > 0.69) as well as the calibrated contrast of the backward mask (P > 0.38). This rules out the possibility that the aforementioned correlations between the brain anatomical differences and the fearful face detection criteria are explained by the differences in the stimulus properties determined through calibration.

**Metacognitive performance**

In addition to the bias at the level of perceptual decision-making (i.e. criteria) as described earlier, we examined potential biases in fearful face perception at the level of metacognition. We analyzed the confidence rating as well as the metacognitive sensitivity which was measured as response-specific meta-d’ for the response of perceiving a fearful face as well as for the response of perceiving a non-fearful (happy/angry) face (Maniscalco and Lau, 2014). Meta-d’ is a signal detection theoretic measurement of metacognitive sensitivity which quantifies how well confidence rating can track the perceptual accuracy (Maniscalco and Lau, 2012), such that meta-d’ becomes larger if confidence is rated higher after a correct than incorrect perceptual response. Although meta-d’ quantifies the metacognitive sensitivity independently from perceptual sensitivity (d’), meta-d’ is calculated on the same scale as d’, enabling an intuitive interpretation of meta-d’ in relation to d’ (Maniscalco and Lau, 2012). Because analysis of meta-d’ is not straightforward when the variances between the two stimulus distributions are unequal (Maniscalco and Lau, 2014), we focused on the
Discrimination Tasks where the equal variance assumption held, as reported earlier.

First, as for confidence rating, a one-way ANOVA with a within-subject factor of Responded emotion revealed a significant main effect \( F(2, 39) = 3.48, P = 0.04, \text{partial } \eta^2 = 0.15 \). This main effect was due to that confidence rating was higher when participants responded perceiving a fearful than angry or happy face, although the differences were only marginally significant \((P = 0.07, P = 0.08, \text{respectively})\) (Figure 5A, left).

This elevated confidence for fearful face perception depended on the perceptual accuracy (Figure 5A, right), as revealed by a significant interaction between Perceptual Accuracy (correct/incorrect) and Responded emotion (fearful/angry/happy) in a repeated measures ANOVA \( F(2, 39) = 12.92, P < 0.001 \). Participants reported higher confidence only when they accurately perceived a fearful face (i.e. report perceiving a fearful face when such face was actually presented) than when they accurately perceived an angry face \((P = 0.002, \text{Bonferroni corrected})\). They also reported higher confidence upon accurate perception of a fearful face when compared with a happy face, yet this difference was not significant \((P = 0.29)\). There was no such difference in confidence rating when perception was incorrect. This interaction suggests that, when a fearful face was perceived compared with a non-fearful angry face was perceived, metacognitive sensitivity was higher in that confidence rating better tracked the perceptual accuracy. Accordingly, meta-\(d'\) significantly differed across three emotions \(A\) main effect of Emotion in one-way ANOVA, \(F(2, 39) = 3.26, P = 0.049, \text{partial } \eta^2 = 0.14\)\), which was due to that meta-\(d'\) tended to be higher for the response of fearful face perception than the response of angry face perception (Figure 5B) \((P = 0.073)\).

**Discussion**

It has been shown that fearful face perception is prioritized in that such faces are often perceived more easily or quickly (Milders et al., 2006; Phelps et al., 2006; Yang et al., 2007; Stein et al., 2009, 2010, 2014; Amting et al., 2010; Stienen and de Gelder, 2011). Yet, it was unclear whether such prioritized perception of fearful faces was merely due to bottom-up processing of low-level physical features (Yang et al., 2007). Alternatively, it could
The authors thank Ellen Tedeschi for assisting the collection of brain images and Brian Maniscalco for advising on the analyses of metacognitive performance.

**Funding**

A.K. was supported by US-Japan Brain Research Cooperation Program (National Institute for Physiological Sciences, Japan) during the early phase of project, as well as by Japan Society for the Promotion of Science (JSPS) (Tokyo, Japan) during the late phase of project. This work was supported by the National Institute of Neurological Disorders and...
Supplementary data

Supplementary data are available at SCAN online.

Conflict of interest. None declared.

References


