

## Supplementary Data

**Behavioral Data.** Average percent correct rates in the scanner were 96% and 74%, respectively, for supra- and perithreshold faces, and 94% and 82%, respectively, for supra- and perithreshold houses. Response times after the forced delay were  $0.624 \text{ s} \pm 0.123$  (mean  $\pm$  SD) for suprathreshold faces,  $0.632 \text{ s} \pm 0.123$  for perithreshold faces,  $0.614 \text{ s} \pm 0.125$  for suprathreshold houses, and  $0.632 \text{ s} \pm 0.119$  for perithreshold houses. Response times were slightly longer for perithreshold noise levels than for suprathreshold levels (difference between suprathreshold and perithreshold was 8.4 ms for faces ( $Z = 1.883$ ,  $p = 0.06$ , non-parametric Wilcoxon test) and 18.0 ms for houses ( $Z = 2.432$ ,  $p < 0.05$ ), respectively).

**Control Analyses:** We hypothesized that higher-level cortical regions computing a decision would have to fulfil two conditions: 1) They should show a greater fMRI response during decisions about suprathreshold images of faces and houses than during decisions about perithreshold images of these stimuli; 2) Their activity should be correlated with the difference between the signals in face- and house-responsive regions. If the absolute difference in activation in face- and house-responsive regions would solely depend on the stimulus that is shown one might expect the regressors used to model these two conditions (higher activation for suprathreshold than for perithreshold images, absolute difference signal) to be perfectly correlated and thus a covariation with the difference signal would not yield additional information. To control for the possibility that the two regressors were correlated we computed the correlation coefficients for each subject and found an average correlation between the two regressors of  $r = 0.092 \pm 0.039$  (mean  $\pm$  S.D.), indicating that this was not an issue.

To control for the possibility that BOLD activity in the DLPFC might show a correlation with the maximum activation in face- or house-selective regions alone we performed the following analysis. We first tested whether changes in BOLD activity in the posterior DLPFC correlated with activity from face-responsive and house-responsive regions, respectively. We found the following correlations with BOLD in DLPFC: |FACE-HOUSE|:  $r = 0.507$  ( $p = 0.000$ ), FACE:  $r = 0.366$  ( $p = 0.011$ ), HOUSE:  $r = 0.146$  ( $p = 0.322$ ) (cf. Fig. S3). However, because |FACE-HOUSE| is a complex measure derived from FACE and HOUSE, it may be that the correlation between BOLD activity in DLPFC and the difference signal can be explained by FACE and HOUSE alone. To specifically test this, we analyzed the partial correlation between |FACE-HOUSE| and DLPFC controlling for FACE and HOUSE and found that the correlation remained significant ( $r = 0.301$ ,  $p = 0.038$ ). Additionally, we used stepwise linear regression (Criteria: Probability-of-F-to-enter  $\leq .050$ , Probability-of-F-to-remove  $\geq .100$ ) to test the correlation between changes in BOLD activity in DLPFC with changes in face- and house-responsive regions and found that the only variable fulfilling the inclusion criterion was |FACE-HOUSE| (Significance: |FACE-HOUSE|  $p < 0.0001$ ; FACE  $p = 0.659$ ; HOUSE  $p = 0.558$ ). These results show that the correlation between changes in activity in DLPFC with the difference signal is better than with either the activity from face-responsive regions or house-responsive regions alone.

To control for the possibility that the correlation between BOLD change in DLPFC and the difference signal might be explained by task difficulty we performed the following analysis. We demeaned |FACE-HOUSE| and DLPFC by subtracting the mean signal change per condition averaged across subjects. In other words, referring to Figure S3, we subtracted the respective mean per condition (perithreshold house, suprathreshold house, perithreshold face, suprathreshold face) from each data point. Even after this process |FACE-HOUSE| and DLPFC were still correlated ( $r = 0.308$ ,

$p < 0.033$ ), indicating that the correlation between BOLD change in DLPFC and the difference signal can not simply be explained by task difficulty.