

# Approach-Avoidance Conflict in Major Depressive Disorder: Congruent Neural Findings in Humans and Nonhuman Primates

Maria Ironside, Ken-ichi Amemori, Callie L. McGrath, Mads Lund Pedersen, Min Su Kang, Satoko Amemori, Michael J. Frank, Ann M. Graybiel, and Diego A. Pizzagalli

## ABSTRACT

**BACKGROUND:** Maladaptive approach-avoidance behavior has been implicated in the pathophysiology of major depressive disorder (MDD), but the neural basis of these abnormalities in decision making remains unclear. Capitalizing on recent preclinical findings, we adapted an approach-avoidance conflict task from nonhuman primate research for use in human functional magnetic resonance imaging (fMRI).

**METHODS:** Forty-two female participants, including 18 unmedicated individuals with current MDD (mean age  $25.2 \pm 5.1$  years) and 24 psychiatrically healthy control subjects (mean age  $26.3 \pm 7.6$  years) completed the adapted approach-avoidance task during fMRI. To probe potential mechanistic factors underlying the observed behavioral and fMRI findings and to inform interpretation of putative group differences, we examined electrophysiological data from 2 female *Macaca mulatta* monkeys performing the approach-avoidance conflict task mimicked in the fMRI study.

**RESULTS:** Findings demonstrated congruent neural correlates of approach-avoidance conflict and aversive responsiveness in the anterior cingulate cortex, including the pregenual cortex, of human subjects and macaques (humans:  $p < .05$  whole-brain corrected; macaques:  $p < .05$ ). The MDD group exhibited aberrant task-related activations in the anterior cingulate cortex, prefrontal cortex, and striatum (all  $ps < .05$ ). Neural effects in the MDD group were cross-sectionally associated with stress and depressive symptoms. Importantly, they also prospectively predicted stress at 6-month follow-up (all  $ps < .05$ ).

**CONCLUSIONS:** Findings indicate that there is conservation of anterior cingulate activation across species and that frontal and striatal regions, in unmedicated humans with MDD, are abnormally responsive during cost-benefit decision making. We suggest that these disruptions could be valuable candidates for translational biomarkers.

**Keywords:** Accumbens, Anterior cingulate cortex, Approach-avoidance conflict, fMRI, Major depressive disorder, Primate

<https://doi.org/10.1016/j.biopsych.2019.08.022>

Major depressive disorder (MDD) is a complex condition characterized by multiple abnormalities, including blunted approach and increased avoidance behavior. Decreased approach behavior predicts future depression (1,2), poor treatment outcomes (3–5), and chronicity (6). Similarly, heightened avoidance contributes to the initiation, maintenance, and relapse of MDD (7–13). Despite these findings, little is known about neural mechanisms underlying maladaptive approach-avoidance (Ap-Av) decision making in MDD. Most prior studies focused on approach and avoidance separately; yet, in daily life, decisions are made in conflict situations by balancing rewarding and aversive outcomes. Moreover, prior human studies used paradigms with few correlates in animals. Therefore, developing cross-species comparisons could be important for understanding mechanisms linked to MDD (14–16) and to potentially reduce current setbacks in drug discovery in clinical neuroscience (17). As a first step, we adapted a nonhuman primate (NHP) Ap-Av conflict paradigm (14) for humans. By ensuring functional

equivalency between human and NHP tasks, our goal was to evaluate with more precision mechanisms implicated in dysregulated Ap-Av behaviors in MDD.

In preclinical studies, Ap-Av decision making is instantiated by a cortico-striato-limbic network (16). In rodents, Ap-Av paradigms recruit the striatum, medial prefrontal cortex (MPFC), hippocampus, and amygdala (16,18–21). Notably, aberrant approach behavior emerged in rodents when the MPFC was disconnected from the striatum (20). Chronic stress increased this nonoptimal behavior, as did optogenetic manipulation of the MPFC-striatal pathway, providing causal evidence that an intrastriatal circuit engaged by the MPFC underlies neural processing of Ap-Av decisions (21).

Studies in NHPs complement these findings by highlighting dissociable roles for the dorsolateral prefrontal cortex (DLPFC) and pregenual anterior cingulate cortex (pACC) in Ap-Av behavior, with the pACC preferentially encoding reward and aversiveness, and the DLPFC preferentially encoding low

motivation (14,22). In particular, the pACC and striatum have emerged as key regions for avoidance-related neural activity in NHPs. Specifically, microstimulation in the pACC and the caudate nucleus can increase avoidance, and these effects are blocked by the anxiolytic diazepam (14,23).

In humans, Ap-Av functional magnetic resonance imaging (fMRI) paradigms uncovered conflict-related activation in the pACC, dorsal ACC (dACC), caudate, DLPFC, and insula (15,24–27), and stimulation of the subthalamic nucleus (STN) enhanced avoidance (28). Here, we tested individuals with MDD and healthy control subjects as they were presented with stimuli identical to those used in macaques, which simultaneously, but independently, indicated varied levels of rewarding and aversive outcomes. This protocol created multiple combinations of rewarding and aversive offers and thus multiple levels of conflict. We compared fMRI findings with electrophysiological data acquired in NHPs from regions shown to be implicated in MDD (pACC, DLPFC, striatum). To our knowledge, our study is the first to adapt task design, modeling, and data analysis from a microstimulation-electrophysiological study in NHPs to probe neural circuitry underlying Ap-Av behaviors in MDD. We hypothesize that compared with healthy control subjects, participants with MDD would show reduced activation associated with reward (striatum) and conflict resolution (ACC) but increased activation associated with aversiveness (STN, amygdala). We further hypothesized that these neural abnormalities would correlate with current and future symptoms.

## METHODS AND MATERIALS

### Human Participants

Twenty-one unmedicated female adults with current MDD (MDD group; mean age  $25.2 \pm 5.1$  years) and 35 psychiatrically healthy control female adults (HC group; mean age:  $26.3 \pm 7.6$  years) participated after providing written informed consent to a protocol approved by the Partners Human Research Committee. For details, see the [Supplement](#).

### Procedures

After screening, participants underwent an imaging session, during which they performed a computerized Ap-Av task. After the scan, they rated stimuli for their valence and arousal. Six months later, participants completed a follow-up clinical session and repeated the self-report questionnaires.

### Human Task

The human Ap-Av task (Figure 1A) was adapted from a prior NHP study (14) ([Supplement](#)). In each trial, participants had to decide (using a joystick) whether to approach or avoid an offer. Approach decisions led not only to the receipt of a reward (points), but also to presentation of an aversive picture with a matching aversive sound; avoidance decisions led to no reward and presentation of a neutral picture. The lengths of 2 parametrically varied, horizontal bars denoted the size of the offered reward points and aversiveness of the outcome picture, respectively. The task included 105 trials: 1) approach-reward trials (only reward outcomes;  $n = 15$ ), 2) avoid-threat trials (only aversive outcomes;  $n = 15$ ), and 3) conflict trials (a combination of reward and aversive outcome;  $n = 75$ ).

### Behavioral Data Analysis in Humans

To quantify the influence of reward and aversiveness on choosing to approach or avoid offers, we estimated Bayesian hierarchical logistic regression models using the *brms* package (29) in R, version 3.3.2 (R Foundation for Statistical Computing, Vienna, Austria). We compared models with different transformations of offered reward and aversiveness using the leave-one-out cross-validation method (30). All models were run as hierarchical models simultaneously estimating individual and group parameters. We report effects of MDD on parameters as credibly different when more than 95% of the posterior distribution is above/below zero.

### Human fMRI Data Acquisition, Preprocessing, and Analysis

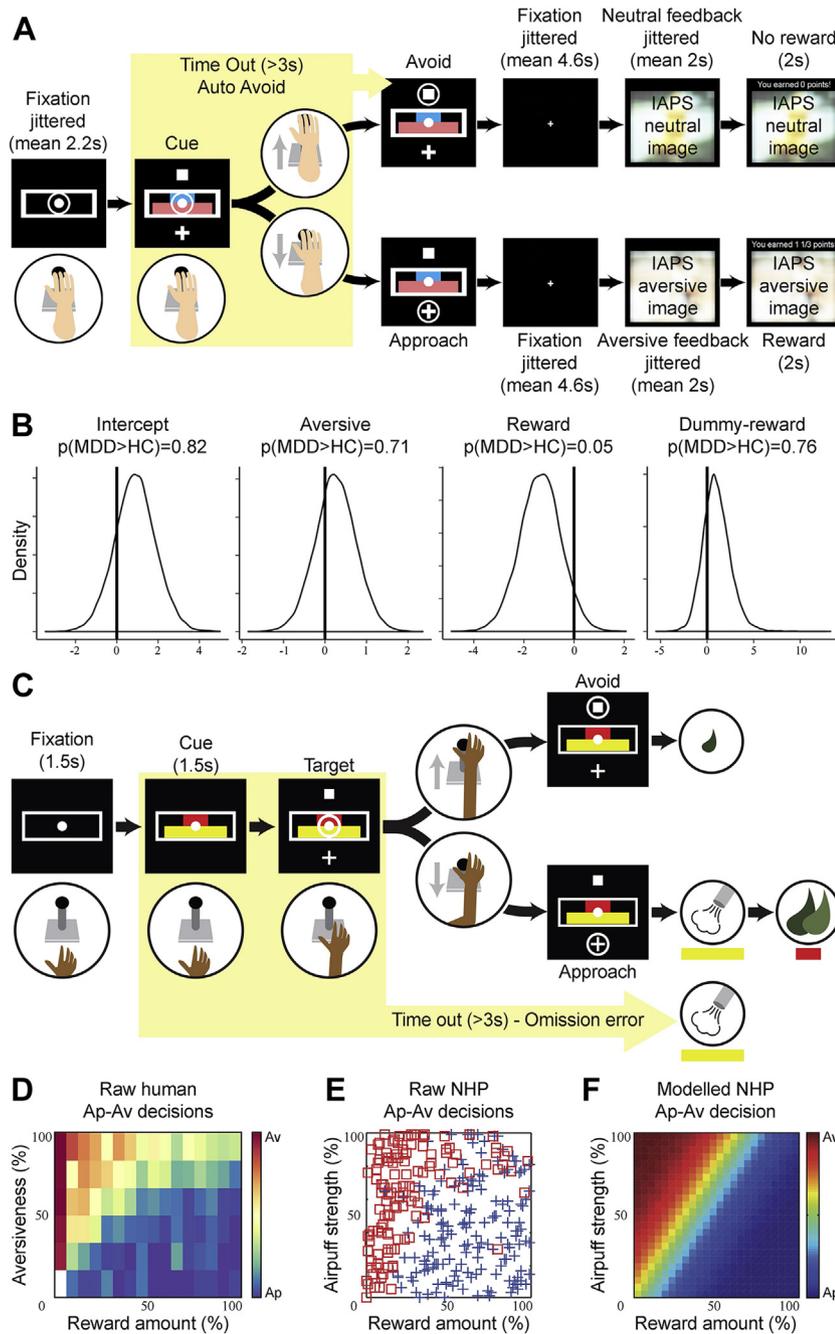
Data acquisition and preprocessing details are provided in the [Supplement](#). The first-level general linear model included 5 regressors (offered choice presentation onsets for approach-reward decisions, avoid-threat decisions, conflict-approach decisions, conflict-avoidance decisions, and feedback). Presentation onsets were also parametrically modulated by trial-by-trial offered reward and aversiveness, and convolved with a hemodynamic response function. For whole-brain analyses, conditions were contrasted to examine 1) approach versus avoidance, averaged across conflict conditions to examine approach or avoidance regardless of conflict, and 2) conflict-approach versus approach-reward to examine the effect of conflict without the potential confound of avoidance activation. For region-of-interest (ROI) analyses, mean activation was extracted from 8 a priori ROIs: bilateral ROIs for the DLPFC (Figure 2A), STN (Figure 2C), NAc (Figure 3A), insula, amygdala, and caudate and single ROIs for the pACC (Figure 2B) and dACC. These values were entered in a mixed-effects linear regression with a between-subjects factor of group (MDD, HC) and within-subjects factors of choice (approach, avoid), conflict (conflict, nonconflict), and, for bilateral ROIs, laterality (left, right). All significant regression interactions were followed up with *t* tests (2-tailed) to examine group differences in approach/avoid and conflict/nonconflict. Effect sizes were estimated using Cohen's *d*. Degrees of freedom differ across contrasts because of bilateral versus unilateral ROIs, outlier exclusions, and use of the Satterthwaite approximation, which considers random effects in mixed-effects models.

### Animal Subjects and Procedure

We studied 2 female *Macaca mulatta* monkeys (monkey A: 7 years of age, 6.8 kg; monkey S: 6 years of age, 7.5 kg) in experiments conducted following the Guide for Care and Use of Laboratory Animals of U.S. National Research Council. All procedures were approved by the Committee on Animal Care of the Massachusetts Institute of Technology. Monkeys were trained to perform an Ap-Av task previously described (14,22,23). For details, see the [Supplement](#).

### NHP Ap-Av Task

As in the human version, in each trial, the monkey had to decide whether to approach or avoid an offer and to indicate her decision by moving a joystick that guided a cursor on a screen (Figure 1C). Two red and yellow horizontal bars



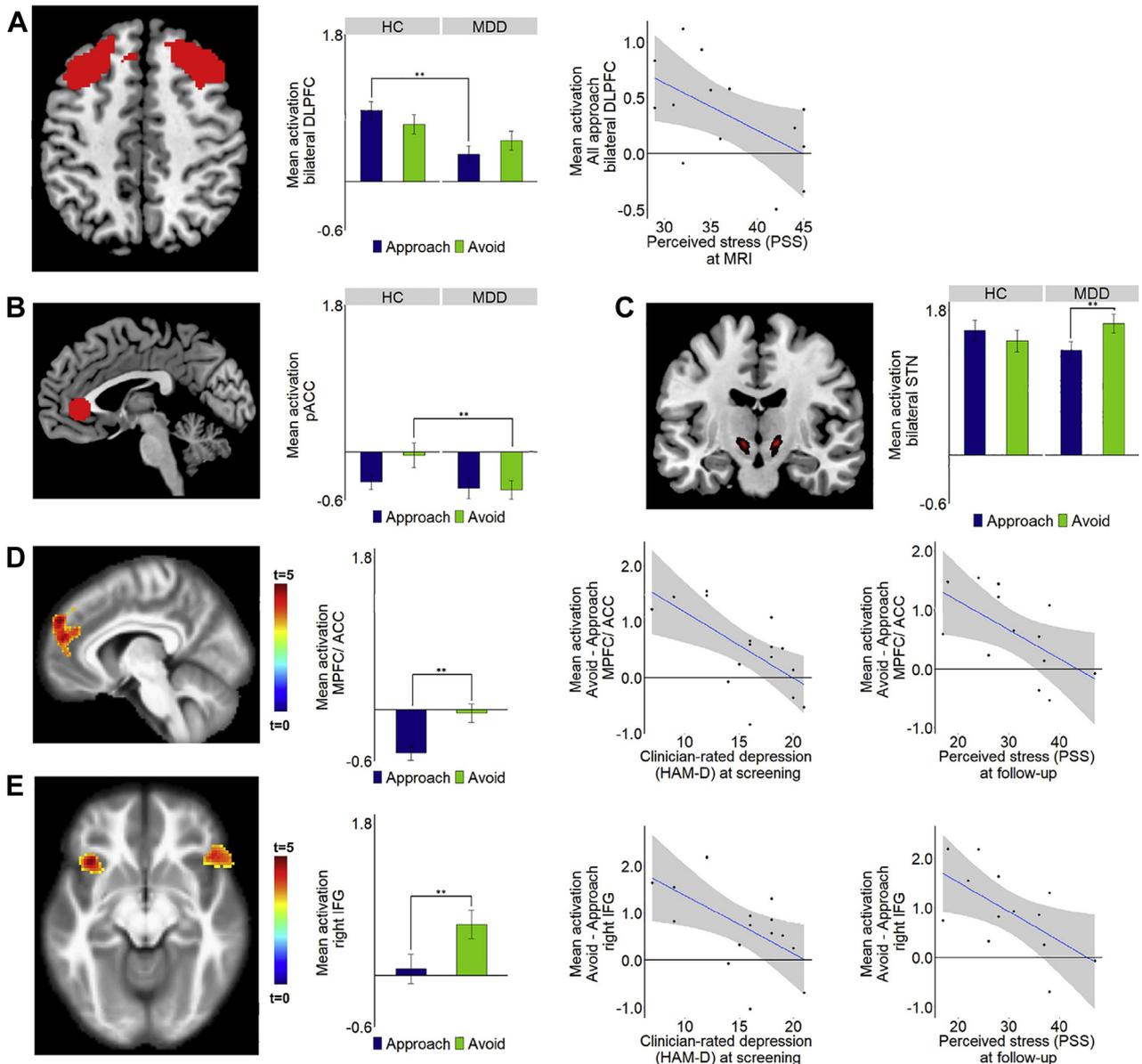
**Figure 1.** Approach vs. avoidance: task and behavior across species. **(A)** Human approach-avoidance (Ap-Av) task: The length of the blue and pink bars indicated the offered points and normative aversiveness of the image/sound presented after approach choice, respectively. The participant could move the joystick to the cross target to indicate an approach choice or to the square target for an avoidance choice. **(B)** Results of Bayesian hierarchical regression: Patients with major depressive disorder (MDD) were less sensitive to reward than healthy control (HC) subjects. **(C)** Nonhuman primate (NHP) Ap-Av task: During the cue period, the red and yellow horizontal bars, respectively, signaling the offered amounts of reward and punishment, appeared on the monitor. The monkeys made a decision between acceptance and rejection of the combined offer and reported this by choosing either of 2 targets (cross for acceptance, square for rejection) that appeared during the response period. **(D)** Raw behavior in the human Ap-Av task. **(E)** Avoidance (red square) and approach (blue cross) decisions made by the monkey in a single session of the Ap-Av task. **(F)** The behavioral model derived by logistic regression with the dataset shown in panel **(E)**. The color scale indicating the probability of choosing avoidance (red) or approach (blue) is shown on the right. IAPS, International Affective Picture System.

appeared on the screen after a 2-second precue period. The lengths of 2 horizontal bars, which were parametrically varied, denoted the offered amount of food reward (red) and the offered pressure of an aversive air puff directed at the monkey's face (yellow). After the 1.5-second cue period, two targets appeared above and below the bars. If the monkey chose the cross target (approach choice), an air puff and food were given at the indicated amounts. When the monkey chose the square target (avoidance choice), no air puff was given, but a

minimal reward was delivered to maintain motivation to perform the task. Target locations were randomly varied.

### Neuronal Recording and Analysis in Macaques

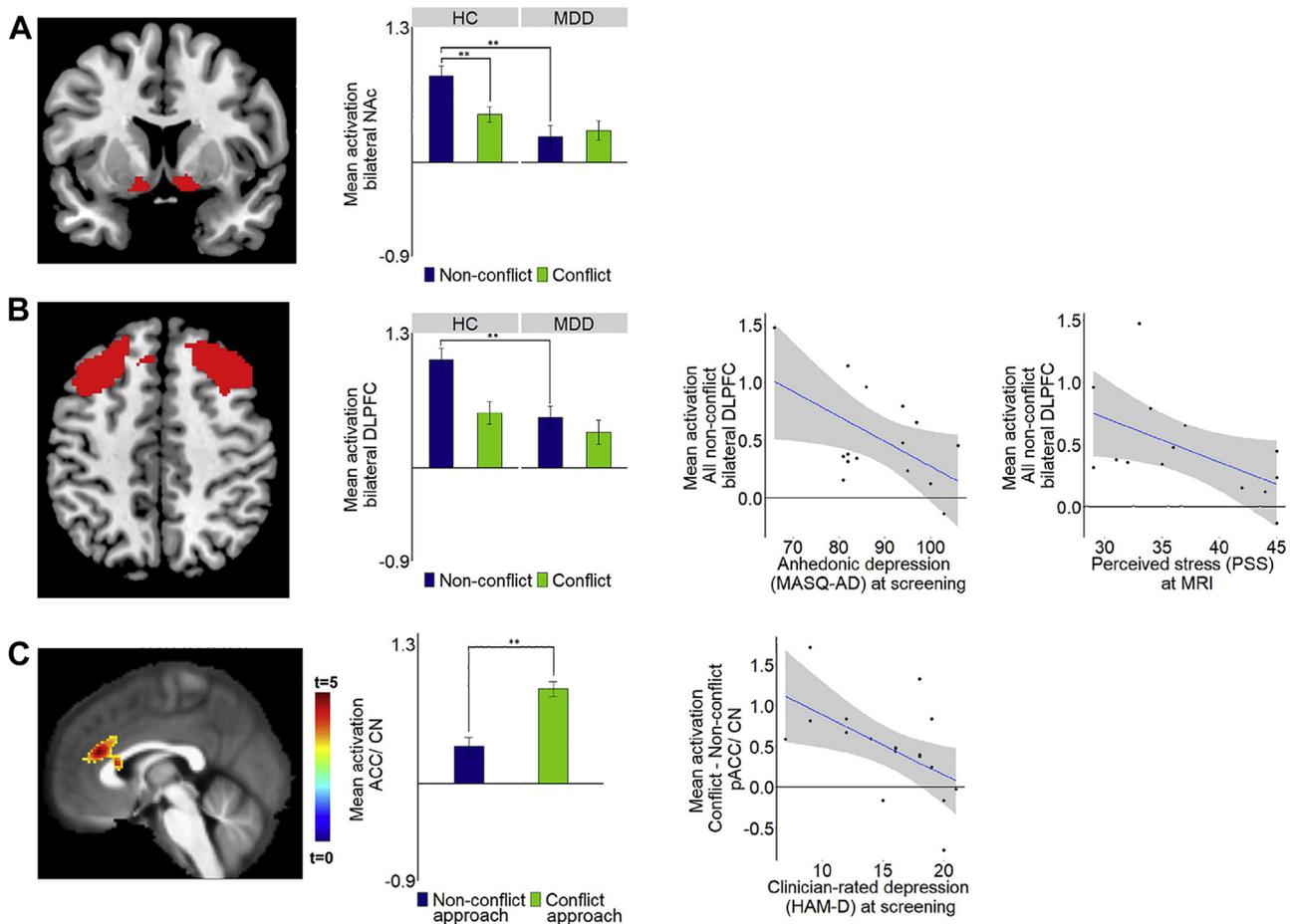
Monkey A had an initial 36 electrodes implanted in neocortical targets (DLPFC:  $n = 24$ ; ACC:  $n = 12$ ), followed by 42 electrodes implanted in a separate session (DLPFC:  $n = 18$ ; ACC:  $n = 24$ ). Monkey S had 12 electrodes first implanted into the ACC, and then, in a separate session, had 30 electrodes



**Figure 2.** Approach vs. avoidance: functional magnetic resonance imaging (MRI) findings. Differences between the major depressive disorder (MDD) and healthy control (HC) groups as well as Pearson correlations between approach- or avoidance-related activation and clinical symptoms are shown for 3 regions of interest: **(A)** bilateral dorsolateral prefrontal cortex (DLPFC), **(B)** pregenual anterior cingulate cortex (pACC) [based on a 10-mm sphere placed on coordinates from a meta-analysis (54)], and **(C)** bilateral subthalamic nucleus (STN). **(D, E)** Thresholded statistical map showing increased activation for avoidance vs. approach trials in the **(D)** medial PFC (MPFC) (superior medial gyrus, superior frontal gyrus) and pACC/dorsal ACC and **(E)** bilateral inferior frontal gyrus (IFG) (see Supplemental Figure S3 for plots and correlations for left IFG). In all these regions, among the MDD group, reduced avoidance-related activation was associated with greater severity of clinician-rated depression (Hamilton Depression Rating Scale [HAM-D]) at screening and higher levels of perceived stress at follow-up. Note: Whole-brain correction performed using a voxel height threshold of  $p < .001$  and a cluster correction threshold of  $p < .05$  (familywise error). Region-of-interest analyses performed using mixed-effects linear regression with a  $p$  threshold of .05 (uncorrected). All follow-up correlations remain significant when controlling for baseline measures with hierarchical regression. See Supplement for full statistics.  $**p < .05$ . PSS, Perceived Stress Scale.

implanted (DLPFC:  $n = 12$ ; ACC:  $n = 18$ ). The DLPFC region targeted corresponded to Walker’s area 46 (31). The ACC consisted of the dACC (areas 8 and 9) and pACC (areas 24 and 32) (32). Data were classified into single-unit activities using Offline Sorter, version 3 (Plexon Inc., Dallas, TX) and analyzed using MATLAB R2018 (The MathWorks, Inc., Natick, MA). To

model parametrically the monkey’s choice pattern, we adopted the econometric conditional logit model (33,34) to infer subjective internal variables. To examine decision-related activity, we analyzed spike activity during the cue period, during which the monkeys had to make a decision, but they did not yet know the direction of joy stick movement required to approach or



**Figure 3.** Conflict vs. nonconflict: functional magnetic resonance imaging (MRI) findings. Differences between the major depressive disorder (MDD) and healthy control (HC) groups as well as Pearson correlations between conflict- or non-conflict-related activation and clinical symptoms are shown for 2 regions of interest: **(A)** bilateral nucleus accumbens (NAc) and **(B)** bilateral dorsolateral prefrontal cortex (DLPFC). **(C)** Thresholded statistical map showing increased activation across all participants for conflict-approach vs. approach-reward trials in the pregenual anterior cingulate cortex (pACC)/dorsal ACC and caudate nucleus (CN). Mean activation extracted and plotted for the cluster shown in panel **(C)**. In the MDD group, reduced conflict-related activation was associated with increased severity of clinician-rated depression (Hamilton Depression Rating Scale [HAM-D]) at screening. Note: Whole-brain correction performed using a voxel height threshold of  $p < .001$  and a cluster correction threshold of  $p = .05$  (familywise error). Region-of-interest analyses performed using mixed-effects linear regression with a  $p$  threshold of .05 (uncorrected). All follow-up correlations remain significant when controlling for baseline measures with hierarchical regression.  $**p < .001$ . See Supplement for full statistics. MASQ-AD, Mood and Anxiety Symptom Questionnaire–Anhedonic Depression; PSS, Perceived Stress Scale.

avoid the offer. To decode neuronal activity during the Ap-Av task, we performed stepwise regression using MATLAB with explanatory variables and added parameters derived from theoretical modeling. Details of neuronal recording, modeling, and statistical analyses are in the Supplement.

## RESULTS

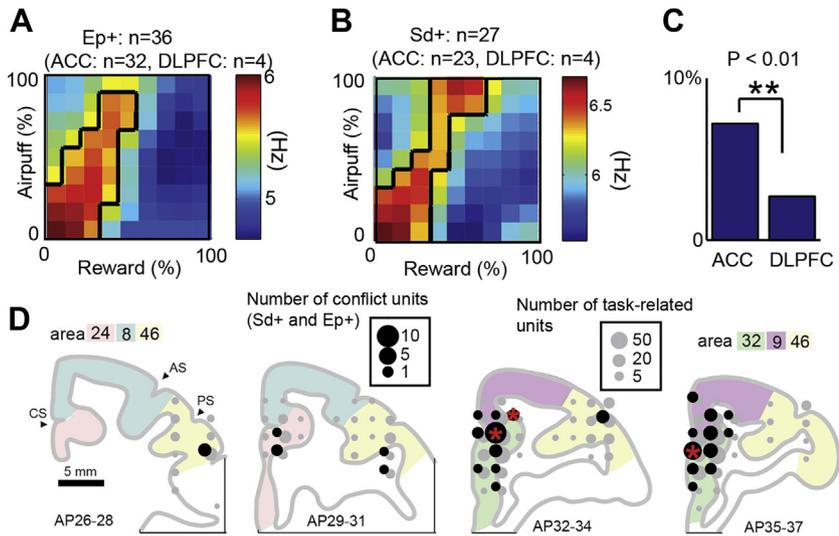
### Human Study

**Reduced Reward Sensitivity in MDD.** Hierarchical Bayesian regression showed that reaction times were significantly increased by conflict ( $\beta = 0.21$ ; 95% confidence interval [CI], 0.14–0.28) and avoidance ( $\beta = 0.14$ ; 95% CI, 0.07–0.21), indicating that the task elicited the expected effects. To investigate the impact of reward and aversiveness on Ap-Av behavior, we compared multiple Bayesian hierarchical logistic regression models. The models differed in the transformations

of reward and aversiveness used to capture observed approach and avoidance (see Supplemental Table S3). The model that best accounted for choice patterns modeled choice (approach = 1, avoid = 0) on trial  $t$  dependent on a logarithmic transformation of the value of offered reward, a direct linear mapping of the value of the offered aversiveness, and a dummy-coded variable for whether offered reward was zero ( $D_{\text{reward}} = 1$ ) or not ( $D_{\text{reward}} = 0$ ):

$$\text{choice}_t \sim \beta_{\text{intercept}} + \log(\text{reward}_t) * \beta_{\text{reward}} + \text{averse}_t * \beta_{\text{averse}} + D_{\text{reward}} * \beta_{D_{\text{reward}}}$$

The analysis confirmed that higher offered reward increased probability to approach ( $\beta_{\text{reward}_{\text{HC}}} = 4.90$ ; 95% CI, 3.79–6.14;  $\beta_{\text{reward}_{\text{MDD}}} = 3.53$ ; 95% CI, 2.38–4.88), whereas stronger aversiveness increased probability to avoid ( $\beta_{\text{averse}_{\text{HC}}} = -2.16$ ; 95% CI, -2.83 to -1.51,  $\beta_{\text{averse}_{\text{MDD}}} = -1.88$ ; 95% CI, -2.62



**Figure 4.** Properties of conflict units. (A, B) Population activity of (A) entropy (Ep+) and (B) standard deviation (Sd+) neurons. (C) Proportion of conflict neurons among all classified neurons in the anterior cingulate cortex (ACC) (cortical areas 8, 24, 32) and bilateral dorsolateral prefrontal cortex (DLPFC) (mainly area 46). We observed conflict units more frequently in the ACC than in the DLPFC ( $*p < .01$ , Fisher's exact test). (D) Distribution of conflict units (Ep+ and Sd+). The size of black and gray circles indicates the number of conflict and task-related neurons at each location, respectively. Red stars on black circles indicate locations in which proportions of conflict units to task-related units were significantly larger than the average ( $*p < .05$ , Fisher's exact test). Note: Units classified with activity correlated positively (+) with standard deviation of decision (Sd+) and entropy of decision (Ep+). AP, anterior-posterior; AS, arcuate sulcus; CS, cingulate sulcus; PS, principal sulcus.

to  $-1.12$ ). The model identified a credible effect of sensitivity to reward across groups: individuals with MDD were less sensitive to reward than healthy control subjects ( $\rho[\beta_{reward_{MDD}} < \beta_{reward_{HC}}] = .05$ ). None of the other coefficients differed between groups (Figure 1B).

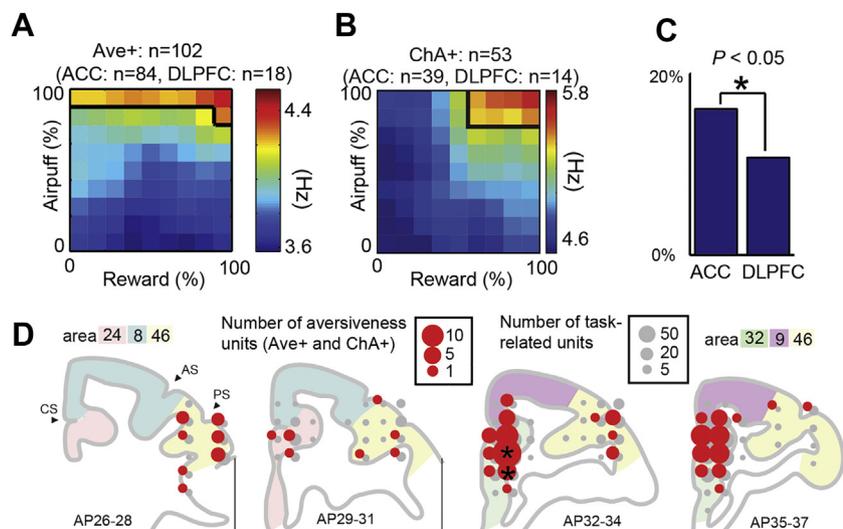
**Imaging Results**

Complete tables of imaging results are presented in Supplemental Tables S4 and S5.

**Aberrant Ap-Av Activation in MDD Is Related to Clinical Symptoms and Stress.** When considering approach versus avoidance (i.e., averaged across conflict conditions), relative to the HC group, the MDD group showed reduced approach-related activation in the DLPFC ( $t_{156} = 3.54$ ,

$p < .001$ ,  $d = -0.55$ ) (Figure 2A) and reduced avoidance-related pACC activation ( $t_{79} = 2.24$ ,  $p = .03$ ,  $d = -0.46$ ) (Figure 2B). Moreover, within-group analyses revealed that the MDD group showed increased avoidance-related STN activation compared with approach-related STN activation ( $t_{138} = -2.06$ ,  $p = .04$ ,  $d = 0.35$ ) (Figure 2C). In the MDD group, reduced approach-related DLPFC activation correlated with higher perceived stress (Perceived Stress Scale) ( $r = -.57$ ,  $p = .04$ ).

Whole-brain-corrected analyses across all participants demonstrated avoidance-related activation in 3 clusters: one cluster in the MPFC and pACC/dACC ( $p < .001$ , whole-brain corrected) (Figure 2D) and bilateral clusters in the inferior frontal gyrus (IFG) ( $ps < .01$ , whole-brain corrected) (Figure 2E). In the MDD group, decreasing levels of avoidance-related activation in these regions were associated with increased baseline depressive symptoms (Hamilton Depression Rating Scale) and



**Figure 5.** Properties of aversiveness neurons. (A, B) Population activity of (A) aversiveness (Ave+) and (B) chosen aversiveness (ChA+) neurons. (C) Proportion of aversiveness neurons among all classified neurons in the anterior cingulate cortex (ACC) and bilateral dorsolateral prefrontal cortex (DLPFC). We observed aversiveness neurons in the ACC significantly more frequently than in the DLPFC ( $*p < .05$ , Fisher's exact test). (D) Distribution of aversiveness neurons (Ave+ and ChA+). The size of red and gray circles indicates the number of aversiveness and task-related neurons at each location, respectively. Black stars on red circles indicate the locations in which proportions of conflict units to task-related units were significantly larger than the average ( $*p < .05$ , Fisher's exact test). Note: Units classified with activity correlated positively (+) with offered aversiveness and chosen aversiveness. AP, anterior-posterior; AS, arcuate sulcus; CS, cingulate sulcus; PS, principal sulcus.

higher perceived stress at follow-up ( $p < .03$ ). Thus, ROI and whole-brain analyses implicated different parts of the ACC (ROI and whole-brain cluster not overlapping) in avoidance behavior, and reduced ACC activation was associated with depressive symptoms and perceived stress.

**Aberrant Conflict-Related Activation in MDD Is Associated With Clinical Symptoms and Stress.** We next probed the effect of conflict by comparing conflict trials (i.e., trials with a combination of reward and aversive offers) and nonconflict trials (i.e., approach-reward or avoid-threat trials). Compared with the HC group, the MDD group showed reduced NAc ( $t_{143} = 4.16, p < .001, d = -0.66$ ) (Figure 3A) and DLPFC ( $t_{155} = 3.62, p < .001, d = -0.56$ ) (Figure 3B) activation during nonconflict trials. In the MDD group, decreasing nonconflict DLPFC activation was associated with higher baseline anhedonia (Mood and Anxiety Symptom Questionnaire–Anhedonic Depression) ( $r = -.54, p = .04$ ) and perceived stress at MRI ( $r = -.54, p = .05$ ). Thus, in ROI analyses, the MDD group showed no differentiation between conflict and nonconflict activations, and decreased DLPFC activation was associated with anhedonia and perceived stress.

Approached conflict was examined by contrasting conflict-approach and approach-reward trials. Using whole-brain-corrected statistics across all participants, we found conflict-related activation in the pACC/dACC and caudate ( $p < .001$ , whole-brain corrected) (Figure 3C). In MDD, decreasing conflict-related activation was associated with greater baseline depression severity ( $r = -.57, p = .02$ ). Thus, whole-brain-corrected analysis implicated the ACC in conflict monitoring, and aberrant conflict-related activation was associated with depressive symptoms.

**Aversiveness Is Tracked by the Human ACC.** Using parametric modulation of all approached offers, we found no group differences in how aversiveness modulated activation in the Ap-Av task ( $p > .05$ ). However, across all participants ( $n = 42$ ), we identified whole-brain-corrected clusters tracking trial-by-trial aversiveness in approach trials in a large cluster in the orbital gyrus, pACC, and dACC and in another cluster in the right IFG and insula (IFG/insula) ( $p < .001$ , whole-brain corrected) (Supplemental Figure S4). These clusters did not survive correction for multiple comparisons in a model comparing reward and aversiveness, and no correlations with clinical measures emerged.

### NHP Study

The goal of the NHP analyses was to address two hypotheses pertinent to the human ACC findings to inform MDD-HC group findings: first, whether the ACC of NHPs contained neurons specifically responding to conflict; and second, whether the ACC contained neurons exhibiting activation related to aversiveness. The data analyzed here were not previously published, except as noted. Most of the methods have been fully described (14,22); accordingly, we describe newly introduced analysis but only summarize other methodology. In total, we isolated 3109 neocortical units from the bilateral DLPFC (mainly cortical area 46) and ACC (areas 8, 9, 24, and 32) of the 2 monkeys, and we classified the units using stepwise regression of 11 explanatory variables (Supplement).

**Conflict Activation in the NHP ACC.** Our human fMRI findings showed activation of the pACC/dACC and caudate associated with conflict in approach decisions (Figure 3C). We thus tested whether the ACC of NHPs contained neurons parametrically responding to conflict. The conflict units consist of 2 groups of units encoding decision-making conflict. We defined the entropy units as those with cue-period activity showing a positive correlation with entropy (Figure 4A) and defined standard deviation units as those with cue-period activity showing a positive correlation with the standard deviation of the Ap-Av choices (Figure 4B). Previous NHP studies have not reported units responding to behavioral conflict in the dACC (35,36) but have reported units responding to behavioral conflict in the DLPFC (37). We thus compared the proportions of the Ap-Av conflict neurons in the ACC and DLPFC. These conflict units were observed significantly more frequently in the ACC than in the DLPFC ( $p < .05$ ) (Figure 4C), suggesting that the ACC contained units with decision-period activity responding specifically for Ap-Av conflict. The distribution of these conflict units was not limited to the pACC (area 32) and was also observed in a broader region including the dACC (area 9) (Figure 4D), resembling conflict activation of the human pACC/dACC (Figure 3C).

The observation of NHP single-unit activity specifically responding to Ap-Av conflict is important, given prior negative results for conflict-specific neuronal responses in the ACC (35–37). Our results clearly show similar neural pACC/dACC responding in humans and NHPs, suggesting a common neuronal mechanism of ACC response to Ap-Av conflict.

**Aversiveness Is Tracked by the NHP ACC.** The human fMRI data indicate greater activation in the pACC and surrounding regions for the degree of aversiveness of the offer (Supplemental Figure S4). We thus tested whether neuronal activity in the macaques exhibited a similar regional bias. Units encoding aversiveness consisted of 2 groups encoding potential and chosen aversiveness. We defined aversiveness units as those with cue-period activity exhibiting positive correlation with the offered air puff (Figure 5A) and chosen aversiveness units as those with cue-period activity showing positive correlation with the size of the air puff to be delivered as a result of the monkey's decision (Figure 5B). These aversiveness units were observed significantly more frequently in the ACC than in the DLPFC ( $p < .05$ ) (Figure 5C). Although these units were found in both the dorsal and ventral banks of the cingulate sulcus, the proportion of the aversiveness units to the task-related units was significantly larger than the average, specifically in the pACC (area 32 or 24) (Figure 5D). These spatial biases in aversiveness unit distribution thus corresponded to the human fMRI data demonstrating pACC activation for aversiveness.

### DISCUSSION

Cross-species models of Ap-Av conflict should be valuable in providing mechanistic information for translational research. For technical reasons, such studies have been lacking. Here, we designed a coordinated study in humans and NHPs with similar experimental protocols in an effort to use the NHP findings to inform interpretation of putative differences in neural activity observed through fMRI of individuals with MDD. Relative to

healthy control subjects, unmedicated participants with MDD exhibited reduced 1) reward sensitivity, 2) ventral striatal and DLPFC activation in nonconflict trials, 3) approach-related DLPFC, and 4) avoidance-related pACC activation. Moreover, unlike healthy control subjects, individuals with MDD showed larger STN activation during avoidance than during approach. These patterns were bolstered by 2 additional sets of findings. First, neural abnormalities during the Ap-Av task were correlated with current stress appraisal and depressive symptoms and predicted stress appraisal 6 months later. Second, across species, conflict and aversiveness were associated with activation in regions of the ACC, validating targets emerging from fMRI analyses. Collectively, findings point to network-level alterations highlighting dysregulation in complex interactions between reward valuation, cost-benefit integration, and conflict resolution.

### Aberrant Reward Sensitivity and Avoidance Signaling in MDD

Individuals with MDD were less sensitive to reward than healthy control subjects. In addition, relative to healthy control subjects, individuals with MDD exhibited reduced pACC activation during avoidance (Figure 2B), suggesting a reduction in normative avoidance activation. Given literature highlighting maladaptive avoidance in MDD (10,13), we speculate that the pACC abnormality might reflect a more automatic (lacking cost-benefit integration) avoidance decision-making style in MDD, potentially reflecting the lack of behavioral sensitivity to reward driving less need for conflict resolution. This hypothesis suggests that maladaptive avoidance in MDD might be linked to abnormalities within network circuitry including regions examined here, the neocortex, the striatum, and the STN. In accord with this speculation, the MDD group showed relatively increased STN activation during avoid decisions (Figure 2C). The STN is proposed to raise decision thresholds in cortico-basal ganglia circuits to prevent approach responses (38), which might be accentuated in MDD. Bilateral STN stimulation induces immobility in the forced swim test in rats (39) and increases avoidance (28) and depressive symptoms (40) in humans.

As hypothesized, and confirming cross-species participation of the ACC and ventrolateral PFC in avoidance behavior and conditioned fear (14,20,22,41–43), whole-brain analyses demonstrated avoidance-related activation in the MPFC/pACC/dACC and bilateral IFG. In individuals with MDD, reduced avoidance-related activation correlated with depression severity (Figure 2D) and predicted higher levels of perceived stress 6 months later. These findings draw parallels to studies in rodent Ap-Av behavior implicating the MPFC-striatal circuit in aberrant valuation of rewards and punishments, demonstrating similar effects of chronic stress and optogenetic inhibition of the medial prefronto-striatal circuit (20,21). In prior analyses of our NHP sample, stimulation of the pACC increased avoidance decisions, and administration of diazepam blocked this effect (14). Future work would benefit from administration of diazepam in humans.

### Blunted DLPFC and NAc Activation in MDD

The DLPFC has been implicated in approach and anticipation of aversiveness in low-conflict decisions (left DLPFC) (24,44) and

avoidance and high-conflict decisions (right DLPFC) (15,45). Electroencephalographic research has linked these asymmetries to MDD (46), but this relationship has not consistently emerged with fMRI. No evidence of laterality emerged, but findings extend earlier reports by showing that MDD is characterized by reduced bilateral DLPFC approach-related activation (Figure 2A). Additionally, MDD was associated with reduced DLPFC activation during nonconflict trials (Figure 3B), and blunted DLPFC activation correlated with increasing anhedonic symptoms and perceived stress. Prior NHP findings demonstrated that activation of DLPFC neurons signaled low motivation (22). Therefore, reduced ability to engage the DLPFC to complete low-conflict/low-motivation trials in MDD represents a potential neural underpinning of impaired anticipation of aversion, stress appraisal, and anhedonia.

Neuroimaging implicates the ventral striatum, particularly the NAc, in the anticipation and valuation of rewards (47,48). Alterations in NAc activations are implicated in a range of psychiatric conditions (49) and are thought to underlie deficits in reinforcement learning and motivation. The reduced NAc activation reported here may thus suggest blunted neural response related to the anticipation and evaluation of reward in a given offer in MDD, an effect that has hitherto not been directly linked to Ap-Av behavior. We found that the MDD group was behaviorally less sensitive to reward. In addition, in the absence of conflict (in approach-reward/avoid-threat trials), the NAc was activated in the HC group but not in the MDD group (Figure 3A). Maladaptive NAc responses to nonconflict choice situations (e.g., easy choices) may be a key underlying feature contributing to impaired approach behavior in MDD.

### Cross-species Function of the ACC

The role of the human ACC in conflict monitoring is well established by prior work in healthy control subjects (15). However, conflict activation in the ACC of NHPs has been debated (Supplemental Discussion). The prior gap between humans and NHPs could stem from differences in task requirements (50) or cognitive demands (51). Here, we focused on Ap-Av conflict in which participants need reconciliation between positive and negative emotional responses. Surprisingly, the 2 prior neuroimaging studies of Ap-Av conflict in MDD reported group differences in multiple regions (including the striatum) but not in the ACC (52,53). Here, we found neural correlates of conflict in pACC/dACC and caudate (Figure 3C). Reduced conflict-related activation, in a region similarly activated by conflict in the NHPs (Figure 4), was associated with depression severity and perceived stress in MDD.

Further cross-species integration stems from the comparison of findings on chosen aversiveness being encoded in the pACC of NHPs. The region identified in the monkeys was in the ventral bank of the cingulate sulcus (posterior part of cortical area 32 and/or anterior part of area 24) (Figure 5). Using parametric modulation, we found that a large region of the ACC (including/adjoining the pACC) also encoded chosen aversiveness in humans (Supplemental Figure S4) (see Supplement for a model comparing reward and aversiveness). Thus, the current NHP and human findings concur in highlighting a role of the ACC in conflict and aversion processing. Given prior NHP findings (14,41), this cross-species integration

should aid future investigations of interventions in humans with MDD and anxiety disorders that could remediate Ap-Av abnormalities.

### Limitations

Despite our integration of behavioral assessments, brain activity measures, and computational modeling, limitations exist. First, behavioral modeling indicated reduced reward sensitivity in MDD, but groups did not differ in avoidance. This pattern points to possible specificity, but participants made significantly more approach decisions than avoid decisions, indicating that the aversiveness of the affective images may have not been potent enough to produce behavioral group differences. Second, whole-brain fMRI analyses were corrected, but the ROI-based regression analyses report corrected and uncorrected statistics. When applying a Bonferroni correction, only NAc and DLPFC group differences remain significant. Also, when comparing aversiveness and reward post hoc in the parametrically modulated findings, aversiveness-related clusters do not survive correction, limiting specificity. Third, although there was high correlation between the normative and subjective ratings for most participants, the assumption that aversive stimuli meant the same to all participants is another limitation. Fourth, the putative interaction between reward valuation and conflict resolution prevented us from separating these 2 aspects of Ap-Av conflict, and it is possible that MDD-related abnormalities in one or both domains might have driven findings. Fifth, because the NHPs did not show depressive-like phenotypes, their data do not immediately inform models of MDD. However, confluence in computational parameters (e.g., avoidance-related pACC activation) modulated by MDD (in humans) and stimulations (in NHPs) allowed us to draw stronger conclusions about fMRI findings in MDD. Future integration of preclinical studies using manipulations relevant to depression (e.g., chronic stress) (19,20) and studies in MDD will be needed for stronger mechanistic models. Finally, only female human participants and macaques were included, limiting generalizability. Despite these limitations, the current cross-species study takes the first steps in developing an Ap-Av conflict model that can be used in humans and NHPs, defining a neural model of avoidance and conflict that correlates with and predicts the symptoms of MDD.

### ACKNOWLEDGMENTS AND DISCLOSURES

The work in humans was supported by National Institute of Mental Health Grant No. R37 MH068376 (to DAP), a Kaplen Fellowship on Depression (to CLM), and a Livingston Fellowship (to CLM), with partial support from the John and Charlene Madison Cassidy Fellowship in Translational Neuroscience (to MI and CLM); the work in NHPs was supported by National Institutes of Health Grant No. R01 NS025529, the CHDI Foundation Grant No. A-5552, Office of Naval Research Grant No. N00014-07-1-0903, Army Research Office Grant No. W911NF-16-1-0474, MEXT KAKENHI Grant Nos. 18H04943 and 18H05131, the Simons Center for the Social Brain, the Naito Foundation, the Uehara Memorial Foundation, and the Saks Kavanaugh Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

We would like to thank Poornima Kumar and Stefanie Nickels for their guidance on the human functional magnetic resonance imaging analyses, Jeffrey Curry for his assistance with NHP analyses, Dan Gibson for his advice on these analyses, Amit Etkin for advice on human anatomical masks, and Malavika Mehta for her assistance with human data collection.

Over the past 3 years, DAP has received consulting fees from Akili Interactive Labs, BlackThorn Therapeutics, Boehringer Ingelheim, and Takeda; and an honorarium from Alkermes for activities unrelated to the current work. No funding from these entities was used to support the current work, and all views expressed are solely those of the authors. All other authors report no biomedical financial interests or potential conflicts of interest.

### ARTICLE INFORMATION

From the Center for Depression, Anxiety and Stress Research (MI, CLM, MSK, DAP), McLean Hospital, Belmont; Department of Psychiatry (MI, DAP), Harvard Medical School, Boston; and McGovern Institute for Brain Research (KA, SA, AMG) and Department of Brain and Cognitive Sciences (KA, SA, AMG), Massachusetts Institute of Technology, Cambridge, Massachusetts; Hakubi Center for Advanced Research (KA), Kyoto University, Kyoto, and Primate Research Institute (KA), Kyoto University, Aichi, Japan; and the Department of Psychiatry and Human Behavior and Department of Cognitive, Linguistic and Psychological Sciences (MLP, MJF), The Robert J. & Nancy D. Carney Institute for Brain Science, Brown University, Providence, Rhode Island.

MI and KA contributed equally to this work.

Address correspondence to Diego A. Pizzagalli, Ph.D., Center for Depression, Anxiety and Stress Research, McLean Hospital, 115 Mill Street, Belmont, MA 02478; E-mail: [dap@mclean.harvard.edu](mailto:dap@mclean.harvard.edu).

Received Apr 5, 2019; revised Aug 8, 2019; accepted Aug 10, 2019.

Supplementary material cited in this article is available online at <https://doi.org/10.1016/j.biopsych.2019.08.022>.

### REFERENCES

1. Wardenaar KJ, Giltay EJ, van Veen T, Zitman FG, Penninx BWJH (2012): Symptom dimensions as predictors of the two-year course of depressive and anxiety disorders. *J Affect Disord* 136:1198–1203.
2. McFarland BR, Shankman SA, Tenke CE, Bruder GE, Klein DN (2006): Behavioral activation system deficits predict the six-month course of depression. *J Affect Disord* 91:229–234.
3. McMakin DL, Olino TM, Porta G, Dietz LJ, Emslie G, Clarke G, *et al.* (2012): Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment resistant depression. *J Am Acad Child Adolesc Psychiatry* 51:404–411.
4. Spijker J, Bijl RV, de Graaf R, Nolen WA (2001): Determinants of poor 1-year outcome of DSM-III-R major depression in the general population: Results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Acta Psychiatr Scand* 103:122–130.
5. Vrieze E, Pizzagalli DA, Demeyttenaere K, Hompes T, Sienaert P, de Boer P, *et al.* (2013): Reduced reward learning predicts outcome in major depressive disorder. *Biol Psychiatry* 73:639–645.
6. Moos RH, Cronkite RC (1999): Symptom-based predictors of a 10-year chronic course of treated depression. *J Nerv Ment Dis* 187:360–368.
7. Ferster CB (1973): A functional analysis of depression. *Am Psychol* 28:857–870.
8. Krantz SE, Moos RH (1988): Risk factors at intake predict non-remission among depressed patients. *J Consult Clin Psychol* 56:863–869.
9. Richter J, Eisemann M, Richter G (2000): Temperament and character during the course of unipolar depression among inpatients. *Eur Arch Psychiatry Clin Neurosci* 250:40–47.
10. Ottenbreit ND, Dobson KS (2004): Avoidance and depression: The construction of the cognitive-behavioral avoidance scale. *Behav Res Ther* 42:293–313.
11. Holahan CJ, Moos RH, Holahan CK, Brennan PL, Schutte KK (2005): Stress generation, avoidance coping, and depressive symptoms: A 10-year model. *J Consult Clin Psychol* 73:658–666.
12. Aldao A, Nolen-Hoeksema S, Schweizer S (2010): Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clin Psychol Rev* 30:217–237.
13. Ottenbreit ND, Dobson KS, Quigley L (2014): An examination of avoidance in major depression in comparison to social anxiety disorder. *Behav Res Ther* 56:82–90.

14. Amemori K, Graybiel AM (2012): Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat Neurosci* 15:776–785.
15. Aupperle RL, Melrose AJ, Francisco A, Paulus MP, Stein MB (2015): Neural substrates of approach-avoidance conflict decision-making. *Hum Brain Mapp* 36:449–462.
16. Kirlic N, Young J, Aupperle RL (2017): Animal to human translational paradigms relevant for approach avoidance conflict decision making. *Behav Res Ther* 96:14–29.
17. Robbins TW (2017): Cross-species studies of cognition relevant to drug discovery: A translational approach. *Br J Pharmacol* 174:3191–3199.
18. Millan MJ (2003): The neurobiology and control of anxious states. *Prog Neurobiol* 70:83–244.
19. File SE, Seth P (2003): A review of 25 years of the social interaction test. *Eur J Pharmacol* 463:35–53.
20. Friedman A, Homma D, Gibb LG, Amemori K, Rubin SJ, Hood AS, *et al.* (2015): A corticostriatal path targeting striosomes controls decision-making under conflict. *Cell* 161:1320–1333.
21. Friedman A, Homma D, Bloem B, Gibb LG, Amemori K, Hu D, *et al.* (2017): Chronic stress alters striosome-circuit dynamics, leading to aberrant decision-making. *Cell* 171:1191–1205.e28.
22. Amemori K, Amemori S, Graybiel AM (2015): Motivation and affective judgments differentially recruit neurons in the primate dorsolateral prefrontal and anterior cingulate cortex. *J Neurosci* 35:1939–1953.
23. Amemori K, Amemori S, Gibson DJ, Graybiel AM (2018): Striatal microstimulation induces persistent and repetitive negative decision-making predicted by striatal beta-band oscillation. *Neuron* 99:829–841.
24. Schlund MW, Brewer AT, Magee SK, Richman DM, Solomon S, Ludlum M, Dymond S (2016): The tipping point: Value differences and parallel dorsal-ventral frontal circuits gating human approach-avoidance behavior. *Neuroimage* 136:94–105.
25. Shenhav A, Straccia MA, Cohen JD, Botvinick MM (2014): Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nat Neurosci* 17:1249–1254.
26. Tom SM, Fox CR, Trepel C, Poldrack RA (2007): The neural basis of loss aversion in decision-making under risk. *Science* 315:515–518.
27. Talmi D, Dayan P, Kiebel SJ, Frith CD, Dolan RJ (2009): How humans integrate the prospects of pain and reward during choice. *J Neurosci* 29:14617–14626.
28. Patel SR, Herrington TM, Sheth SA, Mian M, Bick SK, Yang JC, *et al.* (2018): Intermittent subthalamic nucleus deep brain stimulation induces risk-averse behavior in human subjects. *Elife* 7:e36460.
29. Bürkner P-C (2017): brms: An R package for Bayesian multilevel models using Stan. *J Stat Softw* 80:1.
30. Vehtari A, Gelman A, Gabry J (2017): Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27:1413–1432.
31. Walker AE (1940): A cytoarchitectural study of the prefrontal area of the macaque monkey. *J Comp Neurol* 73:59–86.
32. Saleem KS, Logothetis NK (2012): A Combined MRI and Histology Atlas of the Rhesus Monkey Brain In Stereotaxic Coordinates. San Diego, CA: Academic Press.
33. Train K (2003): Discrete Choice Models Using Simulation. Cambridge, UK: Cambridge University Press.
34. McFadden D (1973): Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. *Frontiers in Econometrics*. New York, NY: Academic Press, 105–142.
35. Nakamura K, Roesch MR, Olson CR (2005): Neuronal activity in macaque SEF and ACC during performance of tasks involving conflict. *J Neurophysiol* 93:884–908.
36. Ito S, Stuphorn V, Brown JW, Schall JD (2003): Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* 302:120–122.
37. Mansouri FA, Buckley MJ, Tanaka K (2007): Mnemonic function of the dorsolateral prefrontal cortex in conflict-induced behavioral adjustment. *Science* 318:987–990.
38. Frank MJ (2006): Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw* 19:1120–1136.
39. Temel Y, Boothman LJ, Blokland A, Magill PJ, Steinbusch HWM, Visser-Vandewalle V, Sharp T (2007): Inhibition of 5-HT neuron activity and induction of depressive-like behavior by high-frequency stimulation of the subthalamic nucleus. *Proc Natl Acad Sci U S A* 104:17087–17092.
40. Strutt AM, Simpson R, Jankovic J, York MK (2012): Changes in cognitive-emotional and physiological symptoms of depression following STN-DBS for the treatment of Parkinson's disease. *Eur J Neurol* 19:121–127.
41. Clarke HF, Horst NK, Roberts AC (2015): Regional inactivations of primate ventral prefrontal cortex reveal two distinct mechanisms underlying negative bias in decision making. *Proc Natl Acad Sci U S A* 112:4176–4181.
42. Wallis CU, Cockcroft GJ, Cardinal RN, Roberts AC, Clarke HF (2019): Hippocampal interaction with Area 25, but not Area 32, regulates marmoset approach-avoidance behavior. *Cereb Cortex* 29:4818–4830.
43. Agustin-Pavon C, Braesicke K, Shiba Y, Santangelo AM, Mikheenko Y, Cockroft G, *et al.* (2012): Lesions of ventrolateral prefrontal or anterior orbitofrontal cortex in primates heighten negative emotion. *Biol Psychiatry* 72:266–272.
44. Spielberg JM, Miller GA, Warren SL, Engels AS, Crocker LD, Banich MT, *et al.* (2012): A brain network instantiating approach and avoidance motivation. *Psychophysiology* 49:1200–1214.
45. Chryssikou EG, Gorey C, Aupperle RL (2017): Anodal transcranial direct current stimulation over right dorsolateral prefrontal cortex alters decision making during approach-avoidance conflict. *Soc Cogn Affect Neurosci* 12:468–475.
46. Davidson RJ (1998): Anterior electrophysiological asymmetries, emotion, and depression: Conceptual and methodological conundrums. *Psychophysiology* 35:607–614.
47. Schultz W (2000): Multiple reward signals in the brain. *Nat Rev Neurosci* 1:199–207.
48. Haber SN, Knutson B (2009): The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology* 35:4–26.
49. Whitton AE, Treadway MT, Pizzagalli DA (2015): Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Curr Opin Psychiatry* 28:7–12.
50. Schall JD, Emeric EE (2010): Conflict in cingulate cortex function between humans and macaque monkeys: More apparent than real. *Brain Behav Evol* 75:237–238.
51. Boschin EA, Brkic MM, Simons JS, Buckley MJ (2016): Distinct roles for the anterior cingulate and dorsolateral prefrontal cortices during conflict between abstract rules. *Cereb Cortex* 27:34–45.
52. Derntl B, Seidel E-M, Eickhoff SB, Kellermann T, Gur RC, Schneider F, Habel U (2011): Neural correlates of social approach and withdrawal in patients with major depression. *Soc Neurosci* 6:482–501.
53. Chandrasekhar Pammi VS, Pillai Geethabhavan Rajesh P, Kesavadas C, Rappai Mary P, Seema S, Radhakrishnan A, Sitaram R (2015): Neural loss aversion differences between depression patients and healthy individuals: A functional MRI investigation. *Neuroradiol J* 28:97–105.
54. Marusak HA, Thomason ME, Peters C, Zundel C, Elrahal F, Rabinak CA (2016): You say 'prefrontal cortex' and I say 'anterior cingulate': Meta-analysis of spatial overlap in amygdala-to-prefrontal connectivity and internalizing symptomatology. *Transl Psychiatry* 6:e944.

# **Approach-Avoidance Conflict in Major Depressive Disorder: Congruent Neural Findings in Humans and Nonhuman Primates**

## ***Supplemental Information***

### **Supplemental Methods and Materials**

#### **Human Procedure**

Participants were right-handed, reported no medical or neurological illnesses and no current use of psychotropic medications. Healthy controls reported no current or past psychopathology. All participants were assessed by a clinician using the Structured Clinical Interview for the DSM-IV (SCID; (1)) and the Hamilton Depression Rating Scale (HAM-D, (2)). Six healthy control (HC) participants and one participant with major depressive disorder (MDD) did not complete the study. Two additional MDD participants were excluded from analyses because their diagnosis was later found to be unreliable. Two HC participants were excluded because of a technical issue with registering their task responses. Three additional HC participants were excluded as their task performance was unreliable. This brought the final sample available for analyses to 18 unmedicated individuals with current MDD diagnosis and 24 psychiatrically healthy individuals. Participants were compensated \$15/hour for their time, \$50 for the MRI session and a \$50 completion bonus.

During the initial screening visit, after the SCID, participants completed a number of self-report questionnaires, including the Beck Depression Inventory (BDI-II; (3)), Snaith Hamilton Pleasure Scale (SHPS; (4)), the Cognitive-Behavioral Avoidance Scale (CBAS; (5)), the Mood and Anxiety Symptom Questionnaire (MASQ; (6)) and the Perceived Stress Scale (PSS; (7)) in order to assess, respectively, depressive symptoms, anhedonic symptoms, behavioral avoidance, subtypes of depressive and anxious symptoms, and perceived stress.

### **Human Approach-Avoidance Task Procedures**

An Ap-Av task was adapted from the prior NHP study (8) (main text Fig. 1A). For each trial, two abutting horizontal bars, one pink and one blue, appeared on the projected computer screen. The length of bars indicated the size of the offers. The blue bar represented the offered number of reward points, which ranged from 0 to 5. The length of the adjoining pink bar indicated the offered normative negative valence rating for a picture from the International Affective Picture Series (IAPS, (9)), which also ranged from 0 to 5. These two offered outcomes were delivered only when the participant made an approach decision. At the same time as the compound visual cue, two target cues (a white cross and a white square) appeared, at randomly programmed positions above and below the cues. A cursor (white circle) controlled by a joystick appeared at the center of the screen. The cues and targets remained on until a response was made. To indicate their decisions, participants used a joystick to move the cursor toward one of the two targets

within 3-s (response period). If the participant did not respond within the allotted 3-s response period, the trial was counted as an auto-avoid and was not included in the analysis. When the participant chose the cross target (approach trial), a fixation cross was presented for a variable period of time (jittered, mean duration 4.6 s). Subsequently, an IAPS picture and matched sound with a normative aversiveness rating of the pink bar in the offer was shown (jittered, mean duration 2-s). This remained on the screen when the reward points with the value of the blue bar were shown (additional 2-s duration). If the participant chose the square target (avoidance trial), the same fixation screen was shown but instead was followed by a neutral IAPS image and sound (jittered, mean duration 2-s). This image remained on the screen when subsequently zero reward points were reported on the screen (additional 2-s duration). To avoid low levels of avoidance, reward points were not associated with any financial reward. After each trial, a variable intertrial interval (jittered, mean duration 2.2 s) occurred. The task included a total of 105 such trials, separated into three runs of 35 trials with a short break in-between, for a total length of ~15 min. There were three trial types: 1) trials with a combination of reward and aversive outcome (conflict-trials, 75 trials); 2) trials with only a reward and no aversive outcome (approach-reward trials, 15 trials), and 3) trials with only aversive outcome and no reward (avoid-threat trials, 15 trials). After finishing the fMRI scan, participants reported their ratings of the valence and arousal of the images to ensure appropriateness of normative ratings as a measure of aversiveness. Piloting of the task in a community sample ( $n = 50$ , 15 males) revealed a sex-specific effect, with males showing less responsivity to potential rewards and

aversion. Therefore, for the fMRI study only females were recruited, which also matched the macaque sample.

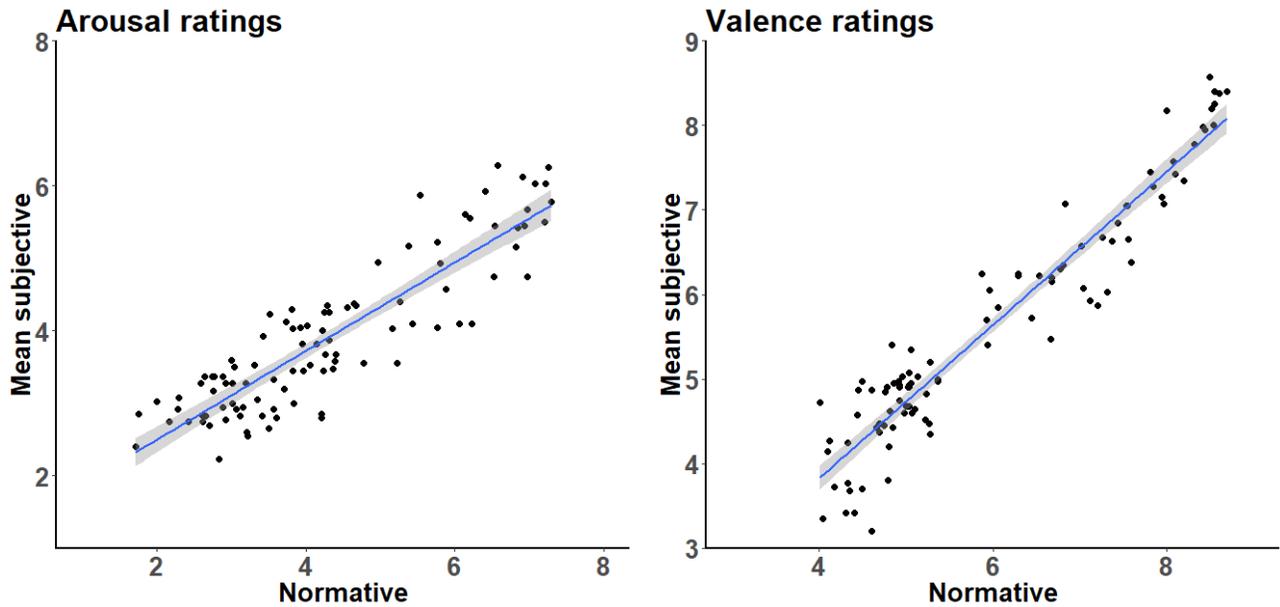
### **Human fMRI Data Acquisition**

A 3T Tim Trio Siemens scanner (Siemens Medical Systems, Iselin, N.J.) equipped with a 32-channel head coil was used to acquire the high-resolution functional and structural MRI data. High-resolution structural data were acquired with a T1-weighted magnetization-prepared rapid acquisition having gradient multi-echo (MPRAGE) imaging sequences with the following acquisition parameters: repetition time (TR) = 2530 ms; echo times (TE) = 3.31, 6.99, 8.85 and 10.71 ms; field of view = 256 mm; voxel dimensions = 1.3 x 1.0 x 1.3 mm<sup>3</sup>; 128 slices. Functional MRI data were acquired using a gradient echo T2\*-weighted echo planar imaging sequence (Connectome sequence, (10, 11)) with the following acquisition parameters: repetition time (TR) = 720 ms; echo time (TE) = 30 ms; field of view = 212 mm; voxel dimension = 2.5 x 2.5 x 2.5 mm; 66 interleaved slices with a multiband acceleration factor of 6 and a GRAPPA acceleration factor of 2.

### **Human fMRI Data Pre-processing**

Functional MRI data were preprocessed and analyzed using Statistical Parametric Mapping software (SPM12; <http://www.fil.ion.ucl.ac.uk/spm>) and statistical software R (12). Distortion correction

was applied using field maps. Functional images were then realigned to the mean image of the series, corrected for motion and slice timing related artifacts, co-registered with the anatomical image, normalized to the 2 x 2 x 2 mm MNI template, and smoothed with an 8mm Gaussian kernel. As described in the main text, the first-level general linear model included five regressors (offered choice presentation onsets for approach reward decisions, avoid threat decisions, conflict approach decisions, conflict avoidance decisions, and feedback). Also, offered presentation onsets for the four decision types were parametrically modulated by trial-by-trial offered reward and aversiveness and convolved with a hemodynamic response function. Normative ratings of aversiveness (13) were used in this analysis to maximize comparability with the NHP work and because the length of the bar representing aversiveness in the human and NHP tasks was based on normative ratings. Note that, as the parametrically modulated time period analyzed was before the image was presented, the length of the aversiveness bar (and thus the normative rating) was the best representation of anticipated aversiveness that each participant could have. Subjective ratings were highly correlated with normative ratings both in terms of valence (Pearson's  $r = 0.95$ ,  $p < 0.001$ ) and arousal ( $r = 0.89$ ,  $p < 0.001$ ) (see *Supplemental Fig. S1* and *Supplemental Tables S1/S2* for individual image ratings). The covariates of no interest included motion realignment parameters and outliers calculated using Artifact Removal Tool (14), and a constant term modelling the baseline of unchanged neural activation was used.



**Supplemental Figure S1: Correlation of normative and subjective International Affective Picture Series (IAPS) ratings:** Normative ratings taken from Lang et al (2008). Subjective ratings carried out by individual participants (mean rating for each image shown).

IAPS image ID	Normative valence rating	Mean subjective valence rating	Normative arousal rating	Mean subjective arousal rating
6150	4.92	4.90	3.22	1.55
3300	7.26	6.68	4.55	3.33
7285	4.33	3.78	3.83	2.45
9090	6.44	5.73	3.97	2.45
7207	4.85	4.43	3.57	1.93
5740	4.79	3.80	2.59	2.28
9041	7.02	6.58	4.64	3.38
7004	4.96	5.03	2.00	2.03
6838	7.55	7.05	5.80	3.93
7035	5.02	4.90	2.66	1.83
2312	6.29	6.25	4.02	3.08
7186	5.37	4.98	3.60	1.80
7170	4.86	4.95	3.21	1.60
6260	7.56	6.65	6.93	4.45
5395	4.66	4.43	4.23	2.45

<b>IAPS image ID</b>	<b>Normative valence rating</b>	<b>Mean subjective valence rating</b>	<b>Normative arousal rating</b>	<b>Mean subjective arousal rating</b>
2490	6.68	6.20	3.95	2.83
2516	5.10	4.65	3.50	1.65
2520	5.87	6.25	4.22	3.00
2870	4.69	4.38	3.01	2.28
3102	8.60	8.38	6.58	5.28
1313	4.35	3.68	4.39	2.58
9921	7.96	7.08	6.52	3.75
7002	5.03	4.68	3.16	1.95
2750	7.44	6.85	4.31	3.25
7100	4.76	4.85	2.89	1.95
9290	7.12	5.93	4.40	2.68
3170	8.54	8.00	7.21	4.50
7180	5.27	4.48	3.43	2.93
9430	7.37	6.63	5.26	3.40
7175	5.13	5.03	1.72	1.40
9404	6.29	6.23	4.67	3.35
7283	4.50	3.70	3.81	3.30
3053	8.69	8.40	6.91	5.13
2850	4.78	4.90	3.00	2.60
9410	8.49	8.58	7.07	5.03
9253	8.00	8.18	5.53	4.88
7140	4.50	4.98	2.92	2.28
9621	6.78	6.30	5.76	3.05
7000	5.00	4.68	2.42	1.75
7185	5.03	5.08	2.64	2.38
9340	7.59	6.38	5.16	3.03
7190	4.45	4.88	3.84	2.00
3530	8.20	7.35	6.82	4.15
7150	5.28	4.35	2.61	1.75
2700	6.81	6.35	4.77	2.55
7233	4.91	4.98	2.77	2.38
9010	6.06	5.85	4.14	2.83
7090	4.81	4.63	2.61	1.83

<b>IAPS image ID</b>	<b>Normative valence rating</b>	<b>Mean subjective valence rating</b>	<b>Normative arousal rating</b>	<b>Mean subjective arousal rating</b>
6350	8.10	7.43	7.29	4.78
2383	5.28	5.20	3.41	1.83
3180	8.08	7.58	5.77	4.23
7495	4.10	4.15	3.82	3.03
3120	8.44	7.95	6.84	4.43
7095	4.01	4.73	4.21	1.85
9046	6.68	6.15	4.31	2.88
3080	8.52	8.20	7.22	5.03
2206	5.94	5.40	3.71	2.20
3130	8.42	7.98	6.97	4.68
8160	4.93	4.75	6.97	3.75
7009	5.07	4.60	3.01	2.00
2570	5.22	4.53	2.76	2.18
7020	5.03	4.90	2.17	1.75
8311	4.12	4.28	3.57	2.33
7820	4.61	4.88	4.21	1.80
9592	6.66	5.48	5.23	2.55
7205	4.44	4.58	2.93	1.78
9910	7.94	7.15	6.20	4.55
7130	5.23	4.83	3.35	2.05
2485	4.31	3.43	3.74	3.13
7010	5.06	4.95	1.76	1.85
2681	5.96	6.05	4.97	3.95
7160	4.98	4.60	3.07	1.93
3064	8.55	8.40	6.41	4.93
9630	7.04	6.08	6.06	3.10
2518	4.33	4.25	3.31	2.53
9080	5.93	5.70	4.36	2.48
6560	7.84	7.28	6.53	4.45
7187	4.93	4.93	2.30	2.08
2381	4.75	4.45	3.04	2.50
9570	8.32	7.78	6.14	4.60
8465	4.04	3.35	3.93	3.05

<b>IAPS image ID</b>	<b>Normative valence rating</b>	<b>Mean subjective valence rating</b>	<b>Normative arousal rating</b>	<b>Mean subjective arousal rating</b>
6242	7.31	6.03	5.43	3.10
5533	4.69	4.48	3.12	1.83
2753	6.83	7.08	4.29	3.35
9280	7.20	5.88	4.26	2.68
9140	7.81	7.45	5.38	4.18
5731	4.61	3.20	2.74	2.38
2722	6.53	6.23	3.52	3.23
7025	5.37	5.00	2.71	1.70
5390	4.41	3.43	2.88	2.38
7235	5.04	4.93	2.83	1.23
3000	8.55	8.25	7.26	5.25
7950	5.06	5.35	2.28	1.93
7380	7.54	7.05	5.88	3.58
5920	4.84	5.40	6.23	3.10
2495	4.78	4.90	3.19	2.28
7351	4.18	3.73	4.25	3.25
2487	4.80	4.20	4.05	2.53

**Supplemental Table S1: Individual normative and subjective International Affective Picture Series (IAPS) ratings:** Normative ratings taken from Lang et al. (2008). Subjective ratings carried out by individual participants (mean rating for each image shown).

<b>Participant ID</b>	<b>Group</b>	<b>Correlation</b>	<b>Participant ID</b>	<b>Group</b>	<b>Correlation</b>
AAC001	HC	0.88	AAC101	MDD	0.86
AAC003	HC	0.89	AAC105	MDD	0.07
AAC004	HC	0.77	AAC106	MDD	0.62
AAC005	HC	0.85	AAC107	MDD	0.86
AAC006	HC	0.46	AAC113	MDD	0.74
AAC009	HC	0.85	AAC115	MDD	0.72
AAC010	HC	0.90	AAC116	MDD	0.76
AAC011	HC	0.84	AAC117	MDD	0.75
AAC015	HC	0.88	AAC120	MDD	0.77

Participant ID	Group	Correlation	Participant ID	Group	Correlation
AAC017	HC	0.89	AAC124	MDD	0.86
AAC018	HC	0.91	AAC126	MDD	0.89
AAC019	HC	0.84	AAC131	MDD	0.87
AAC020	HC	0.88	AAC134	MDD	0.74
AAC021	HC	0.84	AAC136	MDD	0.83
AAC022	HC	0.82	AAC139	MDD	0.73
AAC023	HC	0.88	AAC142	MDD	0.75
AAC029	HC	0.84	AAC143	MDD	0.86
AAC031	HC	0.90			
AAC032	HC	0.82			
AAC033	HC	0.89			
AAC034	HC	0.80			
AAC036	HC	0.45			
AAC039	HC	0.62			

**Supplemental Table S2: Correlation of individual normative and subjective International Affective Picture Series (IAPS) ratings:** Normative ratings taken from Lang et al. (2008). Correlation (Pearson's  $r$ ) of all normative ratings with subjective ratings carried out by individual participants. Subjective ratings were missing for two participants.

### Human fMRI Data Analyses

For whole-brain analyses independent sample and one-sample t-tests were carried out on contrast images with a voxel height threshold of  $p = 0.001$  and a cluster correction threshold of  $p < 0.05$ . Based on prior preclinical and human results obtained with similar constructs/tasks, *a priori* regions of interest (ROIs) were specified bilaterally in the nucleus accumbens (NAc, Pickatlas), amygdala (Oxford-Harvard subcortical atlas, 50% threshold), subthalamic nucleus (STN, FSL subthalamic nucleus atlas, 50% threshold), caudate (Oxford-Harvard subcortical atlas, 50% threshold), insula (Oxford-Harvard

subcortical atlas, 50% threshold) and DLPFC from a frontoparietal control network (15), incorporating the frontal pole, superior frontal gyrus and middle frontal gyrus. Given the key role of the ACC in conflict monitoring and prior NHP work focusing on the pACC, we specified two ACC ROIs, pACC, defined as a single 12mm sphere drawn around coordinates from a meta-analysis (16) and dACC, constituting the remaining portion of the ACC (Oxford-Harvard subcortical atlas, 50% threshold with pACC ROI subtracted), as a single ROI containing dorsal regions. Mean activation during the offer/decision period for each of the four Ap-Av task conditions (approach reward, avoid threat, conflict approach and conflict avoid) was extracted from these ROIs using the SPM12 Summarise function and analyzed with mixed-effects linear regression (degrees of freedom approximated using Satterthwaite's method) using R (12). For bilateral ROIs, mean activation was extracted separately for left and right, thus degrees of freedom differ between bilateral and single region ROIs. Mean activation outliers were identified (less than the first quartile ( $Q1$ ) -  $1.5 * \text{interquartile range (IQR)}$ ) or greater than  $Q3 + 1.5 * \text{IQR}$ ) and were removed before further analysis or correlation with clinical symptoms were performed. Correlations of neural activation with clinical symptoms were also carried out using R and Pearson's correlation coefficients, and hierarchical regression analyses were performed to control for corresponding baseline levels and to test for potential specificity. Follow-up effects from significant interactions were examined with t-tests (2-tailed). For ROI analyses we report both uncorrected and corrected (using the Bonferroni method for 8 ROIs) statistics.

## **Animal Procedure**

### **Animal Approach-Avoidance Task Procedure**

In each trial (main text, Fig. 1C), the monkey was initially presented with a simultaneous white central fixation spot and a gray rectangular frame on a screen. An infrared photobeam sensor detected when the monkey placed its hand on a designated start position. The monkey was trained to hold its hand in the start position for 1.5 s (fixation period). After the fixation, a visual cue, consisting of red and yellow horizontal bars, appeared on the screen. The length of the red bar represented the offered amount of liquefied food, which ranged from 0.1 to 2.0 ml. The length of the yellow bar represented the offered pressure of the airpuff, which ranged from 0 to 50 psi. As in the human version of the task, these two offered outcomes were delivered only when the monkey decided to approach. The length of the bars was variable over 101 steps, independently and randomly. The cues remained on for 1.5 s (cue period), during which the monkey had to maintain start position. If the monkey released the contact (a commission error), the trial was terminated, and an airpuff with pressure indicated by the length of the yellow bar was delivered to the monkey's face. If the monkey continued, after the cue period, two target cues (a white cross and a white square) appeared above and below the cue. At the same time, a cursor (white circle) whose vertical location was controllable by the joystick appeared at the center of the screen. The monkey reported its decision by using a joystick to move the cursor toward one of the two targets within 3-s (response period). The locations of the target cues alternated randomly to be above or below. If the monkey did not respond

within the allotted 3-s response period, the trial was counted as an omission error, and an airpuff as indicated by the yellow bar was delivered. When the monkey chose the square target, the minimum reward (liquefied food, 0.1 ml; corresponding to the minimum offer by the red bar) was delivered to maintain the monkey's motivation. If the monkey chose the cross target, an airpuff whose pressure was represented by the yellow bar was delivered to the monkey's face. Liquefied food, in the quantity signaled by the red bar, was then delivered for 1.5-s.

## **Animal Housing and Neuronal Recording**

### **Animal Subjects and Procedure**

We studied two female *Macaca mulatta* monkeys (monkey A, age: 7 years, 6.8 kg; monkey S, age: 6 years, 7.5 kg) in experiments conducted in accordance with the Guide for Care and Use of Laboratory Animals of United States National Research Council. All procedures were approved by the Committee on Animal Care of the Massachusetts Institute of Technology. The two monkeys were cage-paired with other monkeys and were not housed in isolation. During recording periods, we did not observe abnormal decision-making; such changes were only observed during microstimulation sessions. Before training, we habituated each monkey to sitting in a custom monkey chair. Next, we performed sterile surgery with anesthesia induced by intramuscular atropine (0.04 mg per kg) and ketamine (10 mg per kg) and maintained by inhalation of 1–2.5% sevoflurane with O<sub>2</sub> in order to implant recording chambers

secured by ceramic screws and bone cement. In a subsequent surgical procedure, with the same anesthesia protocol, we implanted electrodes to the regions of interest including the pACC. Post-surgery, the monkeys were kept on analgesics, and prophylactic antibiotics were given intramuscularly both on the day of surgery and thereafter daily for one week. The monkeys were trained to perform an Ap–Av task, previously described in detail (8, 17, 18). We then simultaneously recorded unit activities from the DLPFC and ACC that we used in this study (17). After we started the recording experiments, both monkeys S and A were also used for the microstimulation experiment reported in Amemori & Graybiel (8). Recording and microstimulation experiments were performed in different sessions, and we did not observe abnormal decision-making during the recording sessions. The changes in choice patterns were only observed during microstimulation sessions.

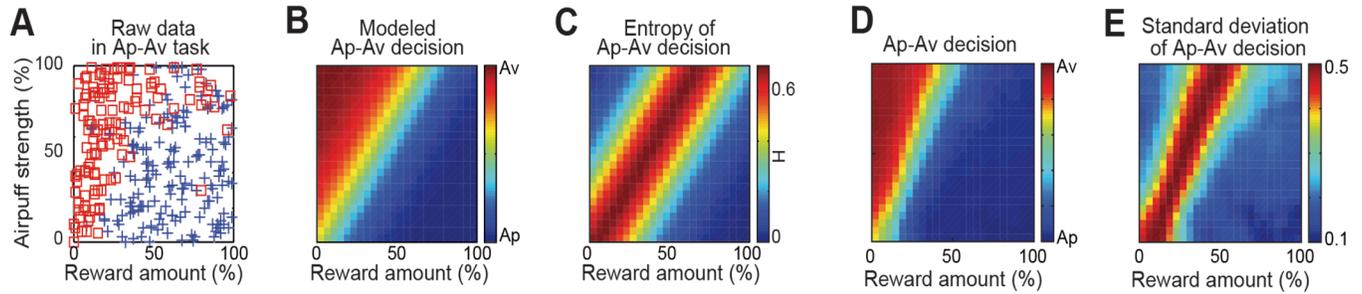
The recording and task control system included five networked computers and other peripheral equipment. Eye positions were monitored by an infrared eye-movement camera system (Eyelink CL; SR Research). Two computers regulated the task based on a CORTEX system developed by the National Institute of Mental Health. For recording, a digital data acquisition system (Digital Lynx; Neuralynx) collected all signals and task event markers. Signals from the microelectrodes were amplified and stored by the Digital Lynx system. Data were later classified into single-unit activities using Offline Sorter (Plexon) and were analyzed using MATLAB (Mathworks).

## Animal Econometrics Models Deriving Entropy

If there were two options associated with the cross and square targets, the probability of choosing the cross target was written as  $p_+ = 1/(1+\exp(-(U_+-U_Y)))$ , where  $U_+$  and  $U_Y$  are the utility of each option (*Supplemental Fig. S2B*). We modeled the difference in utility  $U_+-U_Y$  as  $U_+-U_Y = a_1x+a_2y+a_3$ , where  $x$  was the length of the red bar,  $y$  was the length of the yellow bar, and  $a_1, a_2, a_3$  were determined by the logistic regression. We thus characterized each utility as  $U_+ = U_{Ap} = a_1x+a_2y$ , and  $U_Y = U_{Av} = -a_3$ . The weighted average,  $E_{util} = p_+U_+ + (1-p_+)U_Y$ , is called ‘expected utility’ in economics (19), and is interpreted as corresponding to the ‘chosen value’ (ChV) (20). We consider the chosen value in the Ap-Av task to be enhanced by the expectation of reward and to be suppressed by the expectation of punishment, as this is the parameter that reflects valuation for value-based decision-making. In the Ap-Av task, the entropy of the decisions ( $E_p$ ) (*Supplemental Fig. S2C*) was calculated by the model as  $E_p = -p_+\log p_+ - (1-p_+)\log(1-p_+)$ . Chosen aversiveness (ChA) represents the strength of airpuff that was going to receive in the outcome period. Therefore, the ChA became zero for Av choices, and the strength of airpuff for Ap choices. Standard deviation of choice behavior (Sd) characterized the variability based on the standard deviations directly derived from the averaged behavior.

To examine decision-related activity, we focused on the cue period, during which the monkeys had to make a decision, but during which they did not yet know the direction of movement. To derive the selected combination of explanatory variables by which the explained variable was best fitted to the

neuronal activity, we introduced the *stepwisefit* function of Matlab (Mathworks). The criterion for statistical significance of the F-tests was set at  $p < 0.05$ . To decode neuronal activity during the Ap-Av task, we performed the stepwise regression with the explanatory variables and added parameters derived from theoretical modeling. During the variable selection, we performed multicollinearity checks for each recording session using the Belsley's collinearity test. The combination of variables that exceeded the standard tolerance (a variance decomposition proportion  $>0.5$  and a condition index  $>30$ , established in the *collinest* function of Matlab) was not used. The explanatory variables were the offered reward (Rew, the amount of offered reward indicated by the red bar), the offered airpuff/aversiveness (Ave, the offered pressure of airpuff indicated by the yellow bar), the approach (1) or avoidance (0), the chosen reward (ChR, the amount of reward to obtain), the chosen aversion (ChA, the strength of airpuff to obtain) and the reaction times (RT). We also included the direction of the joystick movement (push or pull) to the explanatory variables. The standard deviation of decisions (*Supplemental Fig. S2E*) and the frequency of omission errors (Om) were derived from the corresponding averaged behavioral parameters. We used the entropy (Ep) and standard deviation (Sd) of decision parameters to characterize the degree of conflict in the Ap-Av task.



**Supplemental Figure S2: Parametric modeling of the decisions by monkey S in a single session in the Ap–Av task.** (A) Avoidance (red square) and approach (blue cross) decisions made by the monkey in a single session of the Ap–Av task. (B) The behavioral model derived by logistic regression with the dataset shown in A. The color scale indicating the probability of choosing avoidance (red) or approach (blue) is shown at the right. (C) The entropy of the decision ( $E_p$ ) of approach (Ap) – avoidance (Av) decisions derived from the model. (D) The mean of choices. (E) The mean of standard deviation of Ap-Av decision.

## Results of Human Study

### Behavioral Results

Model	Reward	Averse	Dreward	Daverse	LOOIC	SE
a	normal	normal	Yes	yes	1761.41	75.42
b	log	normal	Yes	yes	1747.23	75.52
c	normal	log	Yes	yes	1799.11	75.77
d	log	log	Yes	yes	1773.77	76.11
e	normal	normal	No	yes	1988.67	85.49
f	log	normal	No	yes	1839.32	80.56
g	normal	log	No	yes	2052.7	84.51
h	log	log	No	yes	1883.37	81.15
i	normal	normal	Yes	no	1760.37	75.37
j*	log	normal	Yes	no	1745.59	75.05
k	normal	log	Yes	no	1821.21	76.91
l	log	log	Yes	no	1791.87	77.07
m	normal	normal	No	no	1980.02	84.99
n	log	normal	No	no	1834.91	80.14
o	normal	log	No	no	2052.56	84.45
p	log	log	No	no	1891.18	80.59

**Supplemental Table S3. Description and fit of tested models:** 16 models were tested, exhausting all combinations of allowing offered reward and aversiveness to be log transformed or not and to include a dummy coded variable that indicated whether the offered value of reward and aversiveness was 0 ( $D = 1$ ) or not ( $D = 0$ ). Lower values of LOOIC indicate better fit to data. The best fitting model (model j) reported in the results is indicated with an asterisk. LOOIC = Leave-One-Out Cross-Validation information criterion. SE = standard error.

### ***Trial-by Trial Effects***

At the request of an anonymous reviewer we carried out follow-up analyses where we investigated trial-by-trial effects of conflict. To explore this possibility, in the behavioral model we created a regressor that was set to 1 if the previous trial was approached and was high conflict ( $\leq 1$  point (20%) difference between reward and aversiveness). The regressor was estimated to be associated with a slight decrease in probability to approach, but it was not a significant effect for either group ( $\beta_{hc} = -0.36$ , 95% CI = -0.84,0.1,  $\beta_{mdd} = -0.3$ , 95% CI = -0.8,0.2) nor was there an effect of group on the coefficient ( $p(\beta_{hc} > \beta_{mdd}) = 0.438$ ). Further, it did not improve model fit compared to the best-fitting model (LOO-conflict model = 1742.04, LOO-best-fitting model = 1740.61).

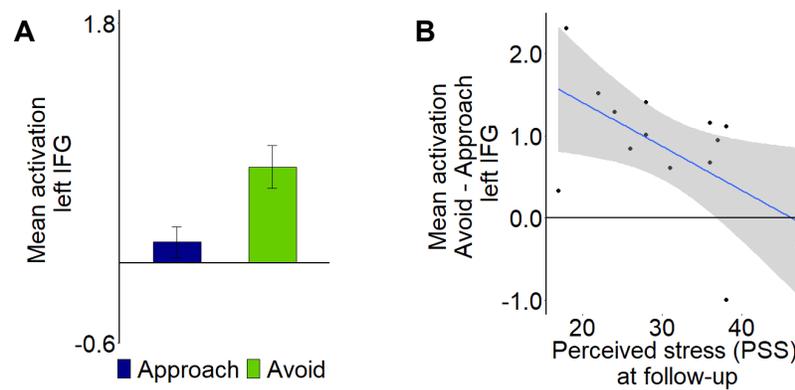
## **Imaging Results**

### ***Approach vs. Avoidance***

When the activation for approach versus avoidance was averaged across conflict and non-conflict trials, whole-brain corrected categorical analyses did not demonstrate any group (MDD vs. HC) differences (all corrected  $ps > 0.05$ ). However, in ROI analyses, mixed-effects linear regression calculations showed that the interaction of *Conflict* (conflict vs. non-conflict) X *Choice* (approach vs. avoidance) X *Group* (MDD vs. HC) was associated with activation in the pregenual anterior cingulate cortex (pACC) ( $\beta = -0.11$ ,  $t(114) = -2.05$ ,  $p = 0.04$ ). Additionally, the *Choice* X *Group* interaction was

associated with activation in the subthalamic nucleus (STN) ( $\beta = -0.10$ ,  $t(275) = -2.12$ ,  $p = .03$ ), and bilaterally in the dorsolateral prefrontal cortex (DLPFC) ( $\beta = 0.075$ ,  $t(267) = -1.99$ ,  $p = 0.05$ ). Follow-up comparisons and correlations with clinical symptoms are reported in the main text results section.

In whole-brain corrected analyses of *all approached vs. all avoided trials* across all subjects ( $n = 42$ ), there was higher activation for avoided trials than approached trials in a large cluster in the medial prefrontal cortex (MPFC) (superior medial gyrus (SMG), superior frontal gyrus (SFG)) and anterior cingulate cortex (ACC) (i.e., MPFC/ACC region) ( $k = 1204$ , XYZ: -12, 56, 34) and in the inferior frontal gyrus (IFG) bilaterally (left:  $k = 261$ , XYZ: 36, 18, -14; right  $k = 334$ , XYZ: 48, 28, -6) (main text, Fig. 2D, E). Within the MDD group, decreased avoidance activation (avoidance - approach) in these clusters (averaged across all voxels shown in Fig. 2D and E, main text) was associated with increased clinician-rated depression severity (Hamilton Depression Rating Scale (HAM-D) scores) at screening (MPFC/ACC:  $r = -0.65$ ,  $p = 0.006$ ,  $n = 16$ ; R IFG:  $r = -0.57$ ,  $p = 0.02$ ,  $n = 18$ ). In addition, decreased avoidance activation in all three clusters was associated with higher levels of self-report perceived stress (assessed by the PSS) in the MDD group at follow-up (MPFC/ACC:  $r = -0.60$ ,  $p = 0.03$ ,  $n = 14$ ; L IFG:  $r = -0.57$ ,  $p = 0.03$ ,  $n = 14$ ; R IFG:  $r = -0.62$ ,  $p = 0.02$ ,  $n = 14$ ), even when controlling for baseline PSS scores (MPFC/ACC:  $\Delta R^2 = 0.28$ ,  $\Delta F(1, 9) = 8.33$ ,  $p = 0.02$ ; L IFG:  $\Delta R^2 = 0.25$ ,  $\Delta F(1, 10) = 8.34$ ,  $p = 0.02$ ; R IFG:  $\Delta R^2 = 0.22$ ,  $\Delta F(1, 9) = 7.38$ ,  $p = 0.02$ ) (main text Fig. 2D, E and *Supplemental Fig. S3*).



**Supplemental Figure S3: Approach vs. avoidance in left IFG.** Refers to main text Fig. 2. (A) Mean activation extracted and plotted for the left IFG. (B) Correlation of left lateralized activation shown in main text Fig. 2E with perceived stress at follow up ( $r = -0.57$ ,  $p = 0.03$ ,  $n = 14$ ).

### ***Conflict vs. Non-conflict***

Whole brain corrected categorical analyses probing conflict did not indicate any group differences (all corrected  $ps > 0.05$ ). However, in ROI analyses, mixed-effects linear regression indicated that the *Conflict X Group* interaction was associated with activation in the bilateral NAc ( $\beta = 0.10$ ,  $t(267) = 2.76$ ,  $p = 0.006$ ). Additionally, the *Conflict X Group* interaction was associated with activation in the bilateral DLPFC ( $\beta = 0.09$ ,  $t(266) = 2.50$ ,  $p = 0.01$ ). Follow-up comparisons and correlations are reported in the main text results section. For correlations with anhedonic depression (assessed by the MASQ-AD subscale), the relationship remained when controlling for anxious arousal (assessed by the MASQ-AA subscale) using hierarchical regression,  $\Delta R^2 = 0.131$ ,  $\Delta F(1,12) = 8.01$ ,  $p = 0.02$ , highlighting specificity.

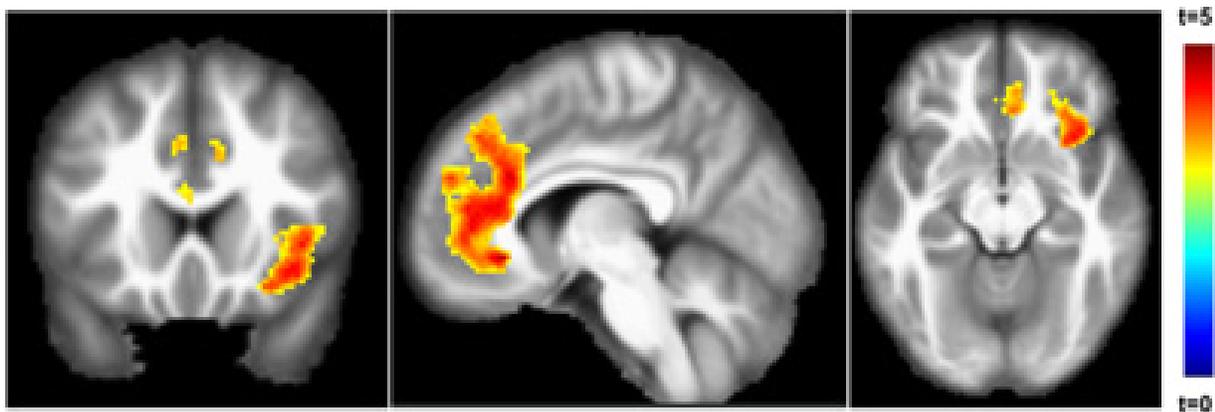
Whole-brain corrected analyses across all subjects ( $N = 42$ ) revealed increased activation for conflict-approach compared to approach-reward trials in a region at the intersection between the pregenual and dorsal ACC (pACC/dACC) leading through the CN into the right middle frontal gyrus ( $k = 1026$ , XYZ: 2, 34, 22) (main text, Fig. 3C). In additional exploratory analyses, reaction time for the conflict and non-conflict conditions was included as a covariate and in these analyses the ACC cluster remained significant. Thus, possible collinearity between reaction times and conflict did not unduly affect the results. Among the MDD group, decreased conflict activation (conflict-approach - approach-reward) in this entire cluster (averaged across all voxels shown in Fig. 3C) was associated with higher clinician-rated depression severity (HAM-D score) at screening,  $r = -0.57$ ,  $p = 0.02$ ,  $n = 18$  (main text, Fig. 3C).

Group effect driving interaction	Condition driving interaction	ROI	t-stat	Significance level
MDD < HC	Approach	DLPFC	3.54	< 0.001 <sup>a</sup>
MDD < HC	Avoid	pACC	2.24	0.03
MDD > HC	Avoid	STN	5.02	0.04
MDD < HC	Non-conflict	NAc	4.16	< 0.001 <sup>a</sup>
MDD < HC	Non-conflict	DLPFC	3.62	< 0.001 <sup>a</sup>

**Supplemental Table S4: Human region of interest fMRI results:** DLPFC: Dorsolateral prefrontal cortex, pACC: pregenual anterior cingulate cortex, STN: Sub-thalamic nucleus, NAc: Nucleus accumbens. <sup>a</sup>Survives a Bonferroni correction for multiple comparisons (8 *a priori* ROIs)

Condition contrast	Cluster peak coordinates	K	Location	t-stat (peak)	Significance level (FWE corrected p-value)
Avoid > Approach	-12, 56, 34	1204	MPFC/ ACC	5.06	< 0.001
Avoid > Approach	-36, 18, -14	261	Left IFG	5.40	< 0.001
Avoid > Approach	48, 28, -6	334	Right IFG	5.02	0.001
Conflict approach > Approach reward	2, 34, 22	1026	ACC/ Caudate	5.47	< 0.001
Aversiveness modulated (all approach trials)	8, 28, 24	2552	Orbital gyrus/ ACC	5.52	< 0.001
Aversiveness modulated (all approach trials)	30, 30, -6	729	Right IFG/ Insula	5.19	< 0.001

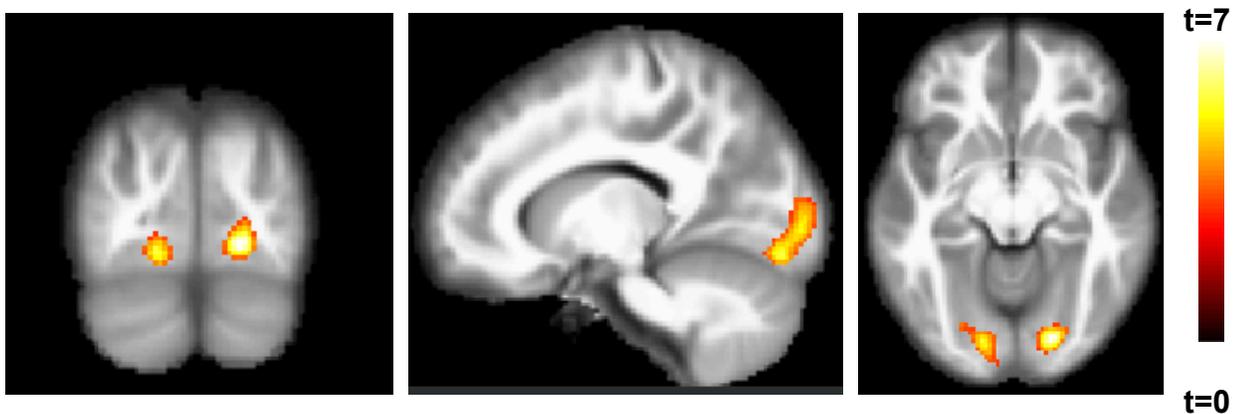
**Supplemental Table S5: Human whole-brain corrected fMRI results:** MPFC: medial prefrontal cortex, IFG: Inferior frontal gyrus, ACC: Anterior cingulate cortex



**Supplemental Figure S4: Regions modulated by aversiveness.** Thresholded statistical map of regions modulated by chosen aversiveness of the offer including the orbital gyrus/pregenual ACC/dorsal ACC and right inferior frontal gyrus (IFG)/insula. Note: Whole-brain correction performed using a voxel height threshold of  $p < 0.001$  and a cluster correction threshold of  $p = 0.05$  FWE.

## Reward

Using trial-by-trial parametric modulation of all offers that were approached, we found no group differences in how reward modulated activation in the Ap-Av task ( $ps > 0.05$ ). Across all participants ( $n = 42$ ), we did not find any clusters in hypothesized regions but did identify whole-brain corrected clusters tracking trial-by-trial reward of the offer in approach trials, specifically a cluster in the right lingual gyrus ( $k = 406$ , XYZ: 18, -86, -8) and left occipital gyrus ( $k = 408$ , XYZ: -14, -88, -12) (*Supplemental Fig. S5*). Upon an anonymous reviewer's request, we additionally examined a model containing both reward and aversiveness as separate parametric modulators and contrasted these parametric effects directly. The parametric effects of aversiveness in the ACC and IFG/insula did not survive correction in this model ( $p > 0.05$ ).



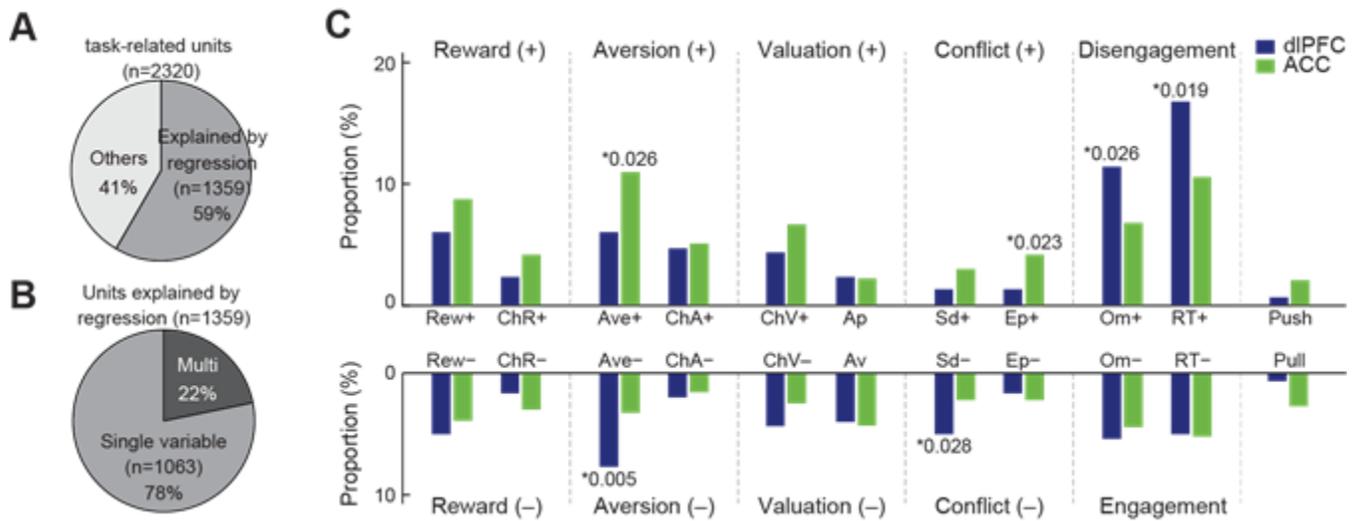
**Supplemental Figure S5: Regions modulated by reward.** Thresholded statistical map of regions modulated by chosen reward of the offer including the right lingual gyrus and left occipital gyrus/cerebellum. Note: Whole-brain correction performed using a voxel height threshold of  $p < 0.001$  and a cluster correction threshold of  $p = 0.05$  FWE.

## Results of Non-Human Primate Study

### Classification of Units

We isolated 3109 neocortical units from the right and left DLPFC (mainly cortical area 46) and ACC (areas 8, 9, 24 and 32) of the two monkeys. Among them, 2320 units were defined as task-related because their firing rates during the cue period activity were significantly different from those during the 1-s time-window before the fixation cue appeared (two sample t-test,  $p < 0.05$ ). We classified those units using the stepwise regression of 11 explanatory variables (see *Supplemental Fig. S6*). 1358 units were explained by the stepwise regression procedure. Importantly, the activities of most units ( $n = 1063$ , 78%) were accounted for by single variables (*Supplemental Fig. S6*). Units were thus classified into the 11 groups of neurons in each region with activity correlated either positively (+) or negatively (-) with offered reward, offered aversiveness, chosen value/utility, choice (approach or avoid), chosen reward, chosen aversiveness, reaction time, standard deviation of decision, entropy of decision, omission, and joystick movement. 117 units (11%) were classified as units encoding conflict (entropy of decision units,  $n=58$ ; standard deviation of decision units,  $n = 59$ ). 221 units (21%) were classified as units encoding aversiveness consisting of offered aversiveness ( $n = 150$ ) and chosen aversiveness ( $n = 71$ ) units (*Supplemental Fig. S6*). Analytical results from neurons coding offered reward, aversiveness, reaction time, utility and omission errors have been reported previously (17). Here, we newly examined neurons

that encoded conflict in decision-making and the offered and chosen aversiveness, to provide analyses parallel to those performed for the data acquired in the human subjects.



**Supplemental Figure S6: Proportion of units coding eleven explanatory variables, (Rew, Ave, ChV, Cho, ChR, ChA, RT, Sd, Ep, Om, and Mv).** (A) Proportion of units explained by stepwise regression. (B) Proportion of units explained by one of eleven explanatory variables. (C) Dissociable properties of DLPFC and ACC neurons. Compared to the ACC (green), the DLPFC (blue) contained significantly higher proportions of neurons encoding RT+, Om+, Ave- and Sd- than the ACC. The ACC contained significantly more proportions of neurons responding to Ave+ and Ep+ than the DLPFC. Significant differences between ACC and DLPFC distributions are indicated by *P*-values (Fisher’s exact test). Data illustrating neurons encoding Rew, Ave, ChV, choice, ChR, ChA, RT+, Om+ and Ave were previously published (17) and are reproduced with permission. Note: Units classified into the 11 groups of neurons in each region with activity correlated either positively (+) or negatively (-) with offered reward (Rew), offered aversiveness (Ave), utility or chosen value (ChV), choice (Ap or Av), chosen reward (ChR), chosen aversiveness (ChA), reaction time (RT), standard deviation of decision (Sd), entropy of decision (Ep), omission (Om), and push/pull joystick movement.

## Supplemental Discussion

The impetus for developing cross-species paradigms comes from current setbacks in drug discovery in clinical neuroscience, which has questioned the utility of animal models of cognitive human disorders (21). Animal models of depression typically encompass measures such as the forced swim test or chronic social defeat, for which there is no direct human equivalent, resulting in poor translation and ultimately, failure of costly phase III trials. In the human study the Ap-Av task was specifically designed to match the non-human primate task as a conscious attempt to bridge this gap. As maladaptive approach-avoidance is a key cognitive phenomenon associated with major depressive disorder, we suggest that showing, for the first time, common neural correlates in a functionally equivalent task represents an important step forward for translational neuroscience specifically addressing current NIMH objectives relating to experimental medicine in psychiatry.

For translational research capitalizing on methods possible in NHPs, it is critically important to determine whether the apparently anatomically corresponding regions in the NHPs and humans have functions within similar domains. For neural mechanisms of conflict resolution in the ACC, the gap between conflict-related neuronal processes in humans and macaques has been emphasized. Cognitive models hypothesize that the human dACC is implicated in cognitive demands that increase the need for behavioral adjustment (22, 23). A series of fMRI studies (24, 25) and single unit recording (26) have supported this view. By contrast, single unit recording in the dACC of NHPs have failed to find conflict-

related neurons responding to behavioral switches (27) and behavioral suppression (28) in conflict conditions. In macaque, conflict-related activity has been reported in the DLPFC, whereas dACC lesions did not induce any deficits in behavioral flexibility (29). This series of studies raised the possibility that the function of the ACC could be different between humans and NHPs (30). By contrast, human fMRI studies reported that cognitive and emotional conflict responses were mediated by different regions of the cingulate cortex (31), and the pACC was implicated in emotional conflict (32). Here, we directly tested these alternative views by examining whether the pACC of humans and NHPs similarly responds to the Ap-Av conflict in which subjects must regulate approach and avoidance motivations. Having established pACC/dACC conflict-related activation in human fMRI (main text, Fig. 3C), we probed NHP data and, in a functionally analogous task, we identified neurons specifically encoding conflict in the pACC/dACC (main text, Fig. 4). Our report is thus the first to show that neurons in the pACC/dACC of macaque monkeys specifically respond to the high conflict conditions in which the animal needed to resolve the conflict between the motivational processes underlying approach and avoidance. In NHPs, conflict resolution of behavioral adjustment has been implicated in the DLPFC (29). By contrast, the proportion of the conflict units that we recorded in the pACC/dACC was significantly larger than in the DLPFC. Our findings thus show that, in NHPs, the pACC/dACC – and less so the DLPFC – is implicated in conflict resolution of approach and avoidance motivations in which positive and negative emotional responses

must be regulated to guide a reasonable value judgment. Such involvement of the pACC/dACC in the regulation of emotional conflict is consistent with previous human fMRI studies (31, 32).

## Supplemental References

1. First MB, Spitzer RL, Gibbon M, Williams JBW (2002): *Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition*. SCID-I/P.
2. Hamilton M (1960): A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 23: 56.
3. Beck AT, Steer RA, Brown GK (1996): Beck Depression Inventory-II (BDI-II). *Psychol Corp*.
4. Snaith RP, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P (1995): A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *Br J Psychiatry*. 167: 99–103.
5. Ottenbreit ND, Dobson KS (2004): Avoidance and depression: the construction of the cognitive-behavioral avoidance scale. *Behav Res Ther*. 42: 293–313.
6. Watson D, Weber K, Assenheimer JS, Clark LA, Strauss ME, McCormick RA (1995): Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *J Abnorm Psychol*. 104: 3–14.
7. Cohen S, Kamarck T, Mermelstein R (1994): *Perceived stress scale - Measuring stress: A guide for health and social scientists*. Oxford University Press New York, NY, 235–283.
8. Amemori K, Graybiel AM (2012): Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat Neurosci*. 15: 776–785.
9. Lang PJ, Bradley MM, Cuthbert BN (1999): *International affective picture system (IAPS): Instruction manual and affective ratings*.
10. Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Glasser MF, *et al.* (2010): Multiplexed Echo Planar Imaging for Sub-Second Whole Brain fMRI and Fast Diffusion Imaging. *PLoS One*. 5: e15710.
11. Xu J, Moeller S, Auerbach EJ, Strupp J, Smith SM, Feinberg DA, *et al.* (2013): Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *Neuroimage*. 83: 991–1001.
12. R Development Core Team (2013): *R: A language and environment for statistical computing*.
13. Lang PJ, Bradley MM, Cuthbert BN (2008): *International affective picture system (IAPS): affective ratings of pictures and instruction manual*. University of Florida, Gainesville.
14. Mazaika PK, Whitfield S, Cooper JC (2005): Detection and repair of transient artifacts in fMRI data. *Neuroimage*. 26: S36.
15. Dosenbach NUF, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RAT, *et al.* (2007): Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci U S A*.

104: 11073–11078.

16. Marusak HA, Thomason ME, Peters C, Zundel C, Elrahal F, Rabinak CA (2016): You say ‘prefrontal cortex’ and I say ‘anterior cingulate’: meta-analysis of spatial overlap in amygdala-to-prefrontal connectivity and internalizing symptomology. *Transl Psychiatry*. 6: e944.
17. Amemori K, Amemori S, Graybiel AM (2015): Motivation and affective judgments differentially recruit neurons in the primate dorsolateral prefrontal and anterior cingulate cortex. *J Neurosci*. 35: 1939–1953.
18. Amemori K, Amemori S, Gibson DJ, Graybiel AM (2018): Striatal microstimulation induces persistent and repetitive negative decision-making predicted by striatal beta-band oscillation. *Neuron*. 99: 829–841.
19. Von Neumann J, Morgenstern O (1947): *Theory of games and economic behavior*, 2nd rev. Princeton University Press.
20. Padoa-Schioppa C, Assad JA (2006): Neurons in the orbitofrontal cortex encode economic value. *Nature*. 441: 223.
21. Robbins TW (2017): Cross-species studies of cognition relevant to drug discovery: a translational approach. *Br J Pharmacol*. 174: 3191–3199.
22. Botvinick MM (2007): Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn Affect Behav Neurosci*. 7: 356–366.
23. Botvinick MM, Cohen JD, Carter CS (2004): Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn Sci*. 8: 539–546.
24. Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD (1999): Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*. 402: 179.
25. Kerns JG, Cohen JD, MacDonald AW, Cho RY, Stenger VA, Carter CS (2004): Anterior cingulate conflict monitoring and adjustments in control. *Science (80- )*. 303: 1023–1026.
26. Sheth SA, Mian MK, Patel SR, Asaad WF, Williams ZM, Dougherty DD, *et al.* (2012): Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature*. 488: 218.
27. Nakamura K, Roesch MR, Olson CR (2005): Neuronal activity in macaque SEF and ACC during performance of tasks involving conflict. *J Neurophysiol*. 93: 884–908.
28. Ito S, Stuphorn V, Brown JW, Schall JD (2003): Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science (80- )*. 302: 120–122.
29. Mansouri FA, Buckley MJ, Tanaka K (2007): Mnemonic function of the dorsolateral prefrontal cortex

- in conflict-induced behavioral adjustment. *Science* (80- ). 318: 987–990.
30. Cole MW, Yeung N, Freiwald WA, Botvinick M (2009): Cingulate cortex: diverging data from humans and monkeys. *Trends Neurosci.* 32: 566–574.
  31. Bush G, Luu P, Posner MI (2000): Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci.* 4: 215–222.
  32. Etkin A, Egner T, Peraza DM, Kandel ER, Hirsch J (2006): Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron.* 51: 871–882.

### KEY RESOURCES TABLE

Resource Type	Specific Reagent or Resource	Source or Reference	Identifiers	Additional Information
Add additional rows as needed for each resource type	Include species and sex when applicable.	Include name of manufacturer, company, repository, individual, or research lab. Include PMID or DOI for references; use "this paper" if new.	Include catalog numbers, stock numbers, database IDs or accession numbers, and/or RRIDs. RRIDs are highly encouraged; search for RRIDs at <a href="https://scicrunch.org/resources">https://scicrunch.org/resources</a> .	Include any additional information or notes if necessary.
Organism/Strain	Rhesus Macaque ( <i>Macaca mulatta</i> )	Primate Center	N/A	
Software; Algorithm	SPM12	The FIL Methods group	RRID: SCR_007_037	
Software; Algorithm	R	The R Project for Statistical Computing	SCR:001905	
Software; Algorithm	Matlab 2017a	Mathworks	RRID: SCR_001622	
Software; Algorithm	bspmview	Spunt, B (2014)	<a href="https://github.com/spunt/bspmview">https://github.com/spunt/bspmview</a>	
Software; Algorithm	CORTEX (VCortex)	National Institute of Mental Health	<a href="https://www.nimh.nih.gov/research/research-conducted-at-nimh/research-areas/clinics-and-labs/in/shn/software-projects.shtml">https://www.nimh.nih.gov/research/research-conducted-at-nimh/research-areas/clinics-and-labs/in/shn/software-projects.shtml</a>	
Software; Algorithm	Offline Sorter	Plexon	<a href="https://plexon.com/products/offline-sorter/">https://plexon.com/products/offline-sorter/</a>	
Software; Algorithm	Chronux	Bokil et al., 2010	<a href="http://chronux.org/">http://chronux.org/</a>	
Other	A 3T Tim Trio Siemens scanner	Siemens Medical Systems		
Other	Connectome EPI sequence	Feinberg et al., 2010; Xu et al., 2013	N/A	
Other	Eyelink 1000	SR Research	Cat#UEPLEDVMCN1X	
Other	Custom built microdrives	Feingold et al., 2012	N/A	
Other	Digital Lynx	Neuralynx	N/A	
Other	Master 8	AMPI	N/A	
Other	A365	WPI	N/A	
Other	L/S Masterflex,	Cole-Parmer	N/A	
Other	20-A	Silent Air	N/A	
Other	900-EIA	Control Air	N/A	