eNeuro Accepted Manuscript

eNeuro

Research Article: Methods/New Tools | Cognition and Behavior

The Neural Basis of Approach-Avoidance Conflict: A Model Based Analysis

https://doi.org/10.1523/ENEURO.0115-19.2019

Cite as: eNeuro 2019; 10.1523/ENEURO.0115-19.2019

Received: 18 March 2019
Revised: 27 June 2019
Accepted: 9 July 2019

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

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

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

eNeuro Accepted Manuscript

Title: The Neural Basis of Approach-Avoidance Conflict: A Model Based Analysis

Abbreviated Title: Model-based analysis of decision conflict

Authors: Samuel Zorowitz[1], Alexander P Rockhill[1], Kristen K Ellard[1], Katherine E Link[1], Todd Herrington[2], Diego A Pizzagalli[3], Alik S Widge[1], Thilo Deckersbach[1], Darin D Dougherty[1]

Affiliations:
[1]Division of Neurotherapeutics, Department of Psychiatry, Massachusetts General Hospital, Charlestown, MA
[2]Department of Neurology, Massachusetts General Hospital, Boston, MA
[3]Department of Psychiatry, McLean Hospital, Belmont, MA

Corresponding Author:
Alexander Rockhill
149 13th St, Room 2625
Charlestown, MA 02129
Email: arockhill@mgh.harvard.edu

Number of pages: 29 pages

Number of figures, tables, multimedia and 3D models (separately):
8 figures (In color)
1 tables
0 multimedia
0 3d models

Number of words for Abstract, Significance Statement, Introduction, and Discussion (separately):
Abstract: 154 words
Significance Statement: 76 words
Introduction: 711 words
Discussion: 2,761 words including citations

Conflict of Interest:
Over the past three years, DAP has received consulting fees or honoraria from Akili Interactive Labs, Alkermes, BlackThorn Therapeutics, Boehringer Ingelheim, Compass, and Takeda, for activities unrelated to the current paper. 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.

Acknowledgements:
This research was sponsored by the U.S. Army Research Office and Defense Advanced Research Projects Agency under Cooperative Agreement Number W911NF-14-2-0045. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the funding sources. Cluster computing resources made possible by Instrumentation Grants 1S10RR023401, 1S10RR019307, and 1S10RR023043, and facilities were funded by grant P41 EB015896. DAP was partially supported by R37 MH068376 and R01 MH101521. KKE was partially supported by NIH National Institute of Neurological Disorders and Stroke (NINDS) Training Program in Recovery and Restoration of CNS Health and Function (T32 NS100663-01).

We thank Mark Helfant for program management and coordination throughout the DARPA SUBNETS effort.

eNeuro Accepted Manuscript

Abstract

Approach-avoidance conflict arises when the drives to pursue reward and avoid harm are incompatible. Previous neuroimaging studies of approach-avoidance conflict have shown large variability in reported neuroanatomical correlates. These prior studies have generally neglected to account for potential sources of variability, such as individual differences in choice preferences and modeling of hemodynamic response during conflict. In the present study, we controlled for these limitations using a hierarchical Bayesian model (HBM). This enabled us to measure participant-specific per-trial estimates of conflict during an approach-avoidance task. We also employed a variable epoch method to identify brain structures specifically sensitive to conflict. In a sample of 28 human participants, we found that only a limited set of brain structures (inferior frontal gyrus, right dorsolateral prefrontal cortex and right pre-supplementary motor area) are specifically correlated with approach-avoidance conflict. These findings suggest that controlling for previous sources of variability increases the specificity of the neuroanatomical correlates of approach-avoidance conflict.

Significance Statement

Approach-avoidance conflict is implicated in many psychiatric syndromes. Previous fMRI studies of this important process have potential biases caused by overlooking individual differences in the evaluation of reward and threat in their analyses. We present a method to model individual differences in approach-avoidance conflict and demonstrate how to incorporate this model into fMRI analyses. We found our approach to have greater specificity than previous studies, which highlights the importance of capturing large variability in participant behavior.

Keywords

Approach-Avoidance Conflict, fMRI, Hierarchical Bayesian Modeling

eNeuro Accepted Manuscript

Introduction

The drive for self-preservation is fundamental to every living organism. Behavioral psychologists have long argued that animals evaluate objects and events in their environments along an appetitive-aversive continuum (Elliot, 2008; Corr, 2013), where animals are motivated to approach things that sustain them (e.g. rewarding or pleasurable stimuli) and to avoid things that threaten them (e.g. harmful or painful stimuli). Approach-avoidance conflict arises in situations where these drives are incompatible, such as when the approach towards reward also increases the possibility of danger. Approach-avoidance conflict is an important phenomenon as it is thought to be core to the etiology and maintenance of psychiatric disorders including depression and anxiety (American Psychiatric Association, 2013).

In recent years, many studies have investigated the neural substrates underlying approach-avoidance conflict using electrophysiology in rodents (Friedman et al., 2015) and non-human primates (Amemori et al., 2015) and neuroimaging in humans (Talmi et al., 2009; Park et al., 2011; Bach et al., 2014; Aupperle et al., 2015; O'Neil et al., 2015; Schlund et al., 2016; Loh et al., 2017). The results of the human neuroimaging literature have implicated a diverse collection of brain structures in approach-avoidance conflict including cortical structures such as the anterior cingulate, insula, orbitofrontal cortex, and dorsolateral prefrontal cortex, and subcortical structures including the amygdala, hippocampus, and striatum. There is considerable heterogeneity in these findings, however, such that none of the aforementioned brain structures are consistently identified as being involved in approach-avoidance conflict across these studies. This naturally prompts the question of where some of the variability might stem from.

One possibility is that the heterogeneity reflects variability in approach-avoidance behavior across participants. Approach-avoidance tendencies are naturally varying across

eNeuro Accepted Manuscript

individuals (Carver & White, 1994), such that there are robust individual differences in the valuation of reward and threat cues. As such, the point of maximal approach-avoidance conflict is unlikely to be the same across participants. Ignoring these individual differences and averaging across them, however, has been shown to reduce contrast statistics in fMRI group level analysis (Ahn et al., 2011). One solution is to explicitly model individual differences in approach-avoidance conflict, such as with hierarchical Bayesian modeling (Kruschke, 2015), and incorporate trial-by-trial estimates of approach-avoidance conflict into the fMRI analysis in order to align participants along a latent evaluation space (O'Doherty et al., 2007; Ahn et al., 2011). In doing so, we are less likely to average out conflict-related changes in BOLD signal.

A second possibility is that the heterogeneity in findings directly reflects variability in previous modeling of conflict-related changes in BOLD signal. A hallmark feature of approach-avoidance conflict is prolonged reaction times. Interpreting changes in BOLD signal between two conditions that also involve differences in response times is challenging, however, due to the time-on-task effect (Taylor et al., 2014). Because the BOLD signal sums approximately linearly as a function of stimulation duration (Dale & Buckner, 1997), brain structures not directly involved in the representation of approach-avoidance conflict may still show increases in BOLD signal by virtue of prolonged processing of the constitutive elements of conflict (e.g. rewarding or threatening stimuli). Controlling for response time is necessary then to identify brain structures that are directly involved in the processing of approach-avoidance conflict (brain regions that show increased intensity of activity, not just prolonged activity). With the exception of Talmi et al. (2009), the neuroimaging studies of approach-avoidance conflict cited above do not document having incorporated response times into their fMRI analyses.

eNeuro Accepted Manuscript

In the present study, we investigated the neural signatures of human approach-avoidance conflict with functional neuroimaging controlling for the issues discussed above. We measured changes in the fMRI BOLD signal as participants completed an approach-avoidance conflict task. In the task, participants repeatedly chose between a risky option, returning greater reward at the risk of potential electrical stimulation, and a safe option, returning a much smaller reward but no risk of electrical stimulation. Using a novel hierarchical Bayesian model (HBM), we estimated participants' per-trial approach-avoidance conflict and used these to inform our fMRI analyses. Moreover, we controlled for the time-on-task effect using the variable epoch method (Grinband et al., 2008) in order to identify brain structures that showed greater intensity of activity, rather than prolonged activity, during approach-avoidance conflict. We found that using these methods increased the specificity of the structures responding to conflict.

Methods

Subjects

Thirty-six individuals (13 females, 23 males, age: M = 33.94 yrs, SD = 8.80) were recruited from the Greater Boston area to participate as healthy volunteers in a research program to develop novel deep brain stimulation (DBS) technologies (Widge et al., 2017). All participants reported being right-handed and without a current or past diagnosis of a psychiatric or neurological disorder and were in the normal healthy range for the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998). Women were scanned at or near the ovulation phase of their menstrual cycles (when estradiol is lowest) to minimize potential gender confounds (Zeidan et al., 2011). The study was approved by the Partners Healthcare System Human Research Committee, and all participants provided written informed consent prior to

eNeuro Accepted Manuscript

enrollment. Participants were paid $600 for the successful completion of the larger study protocol.

Eight individuals were excluded from analysis: five due to technical complications (see below), two for missing responses for more than 20% of trials, and one due to corrupted DICOMs. This resulted in a final sample of 28 participants (10 females, 18 males).

Task

We employed a modified version of the Aversion-Reward Conflict (ARC) task (Sierra-Mercado et al., 2014). During this task, participants make a series of choices between two options: a safe option or a risky option (Figure 1). Selecting the safe option returns a reward of $0.01 and the participant never receives electrical stimulation. In contrast, selecting the risky option returns a reward between $0.05 and $0.95, and the participant receives electrical stimulation with probability 10%, 50% or 90%, as indicated by a bar in the center of the screen. This required participants to evaluate their preference for a greater reward with a risk of electrical stimulation relative to a lesser reward with no risk of electrical stimulation. Participants were instructed to choose as fast as possible without choosing randomly and were informed that their choices would be reflected in their final study payment. (In fact, each participant was compensated with a generous flat payment.) Prior to starting the task, participants were asked to report back the instructions so that their comprehension could be verified. Next, participants completed ten practice trials to become accustomed to the timing of the task.

This ARC task had three levels of risk: 10%, 50%, and 90% likelihood of electrical stimulation. Rewards were sampled from all cent values between $0.05 and $0.95. Trials were counterbalanced such that there were an equal number of trials at each risk level, while rewards were uniformly and equally sampled within each risk level. Each participant completed 108 total

5

eNeuro Accepted Manuscript

trials and the order of trials was kept constant for all the participants. Long inter-trial intervals of 10.5 ± 0.875 seconds separated sequential trials in the task (a slow event-related design). The duration of the full task was 28.5 minutes.

Electrical stimulation was administered to the ankle through a Coulbourn Aversive Finger Stimulator (Harvard Apparatus, E13-22; maximum level of stimulation = 4.0 mA). The amperage of electrical stimulation was calibrated individually for each participant prior to performing the ARC task. Participants experienced increasing levels of stimulation until they reported reaching a subjective threshold qualified as "highly annoying but not painful". For five participants this threshold could not be established because the highest stimulation setting of 4.0 mA was too painful, but penultimate 2.3 mA setting was not experienced as annoying. These participants did not exhibit behavioral variation (i.e., they always accepted the risky choice) and consequently these participants were excluded from the analysis.

Behavioral Analysis

Our aim was to infer the level of approach-avoidance conflict experienced by each participant during every trial. We devised a novel HBM that predicts participants' choices (safe or risky option) and response times. The decision to model response times was motivated by well-documented relationship between decision conflict and prolonged response times and prior demonstrations that including response times in behavioral models improved the accuracy of single-trial parameter estimation (Prerau et al., 2009; Pederson et al., 2016). The model is composed of a logistic regression on the choice data and a gamma regression on the response times. We assume the binary choice responses, y ∈ (0 = safe choice, 1 = risky choice) are drawn from the Bernoulli distribution:

$$p(y_{ij}|\theta_{ij}) = \theta_{ij}^{y_{ij}}(1-\theta_{ij})^{1-y_{ij}}$$

6

eNeuro Accepted Manuscript

where θ_{ij} is the likelihood-of-take for trial *i* and participant *j*, and is itself estimated from:

$$\theta_{ij} = logistic\left(\beta_{0_j} + \Sigma\beta_{nj}x_{nij}\right)$$

Here, β_{0_j} is the intercept for participant *j*; the remaining β_{nj} regression coefficients reflect the modulatory influence of independent variables, *X*, on the baseline likelihood-of-take. In this model, there are three independent variables: 50% risk (β_1), 90% risk (β_2), and reward (β_3). The 50% risk (β_1) and 90% risk (β_2), coefficients are binary predictors, whereas reward (β_3) is a continuous predictor that was normalized to have mean = 0 and standard deviation = 1. The intercept term, β_0, thus reflects the likelihood of take for 10% risk and $0.50 reward offer.

The continuous response times, *z*, are assumed to be drawn from the gamma distribution:

$$p(Z_{ij}|k_j, \mu_{ij}) = Gamma\left(k_j, \frac{k_j}{\mu_{ij}}\right)$$

where k_j is the shape parameter for participant *j* and μ_{ij} is the mean of the distribution predicted by:

$$\mu_{ij} = \alpha_{0j} + \alpha_{1j} \cdot d_{ij}$$

We chose a gamma distribution because it is well-suited for characterizing response times and other strictly positive data with a long rightward tail (Yousefi et al., 2015). Here, α_{0j} was the average response time for participant *j* and α_{1j} was the slope term determining how much response time increased with conflict. We represent conflict, d_{ij}, as the inverse of the distance-to-decision boundary of trial *i* for participant *j*, represented as:

$$d_{ij} = 0.25 - (0.5 - \theta_{ij})^2$$

This measure, *d*, has the shape of an inverted parabola. It is greatest when $\theta = 0.5$, or when a participant is equally likely to select the safe or risky option. It is smallest when $\theta = 0.0$ or $\theta = 1.0$, or when a participant is most likely to select the safe or risky option, respectively.

7

eNeuro Accepted Manuscript

Therefore, *d* reflects the degree of conflict a participant experienced during the evaluation phase of a given trial. The model fit then identified the set of parameters that maximized the joint likelihood of both the choice and response time data due to the relationship between θ and *d*.

As a hierarchical model, each of the participant-level regression parameters defined above (e.g. $\alpha_{0j}, \alpha_{1j}, \beta_{0j}, \beta_{1j}, \ldots, \beta_{nj}$) are drawn a corresponding group-level distribution, centered at group-level means (e.g. $\alpha_{0G}, \alpha_{1G}, \beta_{0G}, \beta_{1G}, \ldots, \beta_{nG}$). Thus, the model simultaneously estimates group- and participant-level parameters, partially pooling the data so as to minimize the influence of outliers. Figure 2 presents a detailed diagram of the model which includes the choice of priors. We assumed Student's t-distribution priors on the choice (β) regression coefficients to ensure robust logistic regression (Gelman et al., 2008; Ghosh et al., 2017) using the recommended degrees of freedom, $\eta = 5$ (Stan Development Team, 2017).

The behavioral model was fit using Hamiltonian Monte Carlo (HMC) sampling in Stan v2.15 (Carpenter et al., 2017) with four chains of 2000 steps each (1000 burn-in, thinning = 4), yielding 1000 posterior samples total. The convergence of the chains was computed using the $\hat{R}$ statistic (Gelman et al., 2014), which measures the degree of variation between chains relative to the variation within chains. The Stan development team recommends as a rule of thumb that all parameters have $\hat{R}$ statistics no greater than 1.1. All parameters in our showed good convergence ($\hat{R} \approx 1$). Similarly, the number of effective samples approached 1000 for most parameters indicating that the chains exhibited low autocorrelation. Once fitted, per-trial estimates of *d* were generated by multiplying the observed trial features (risk level and reward value) by the modal individual-level parameter estimates.

8

eNeuro Accepted Manuscript

Image Acquisition and Preprocessing

All MRI scans were completed at the Athinoula A. Martinos Center for Biomedical Imaging. Of the 28 participants included in this analysis, 20 were scanned using a 3T Siemens Trio scanner and 8 were scanned using a 3T Siemens Prisma scanner (scanner type was entered as a covariate in the analyses). All participants were scanned with a 32-channel head coil. Foam cushions were used to restrict head movements. Task images were projected using a rear projection system and PsychToolbox (V3) stimulus presentation software (Kleiner et al., 2007).

For each participant, both structural and functional images were collected. The structural sequences involved a high-resolution, four-multiecho, T1-weighted, magnetization-prepared, gradient-echo image (TR = 2510 ms, TE = 1.64 ms, flip angle = 7^{o}, voxel size = 1.0 x 1.0 x 1.0 mm) (van der Kouwe et al., 2008). Functional images were acquired using a multiband SMS-3 T2*-weighted echo-planar-imaging (EPI) sequence sensitive to blood-oxygen-level dependent (BOLD) contrast (TR = 1750 ms, TE = 30 ms, flip angle = 75^{o}, voxel size = 2.0 x 2.0 x 2.0 mm, PAT = GRAPPA, accelerated factor TE = 2). Sixty-three interleaved slices were aligned perpendicular to the plane intersecting the anterior and posterior commissures, and the whole brain was imaged (FOV = 220 mm). For the purpose of EPI-dewarping, a fieldmap was also collected for each participant (63 interleaved slices, TR = 500 ms, TE 1 = 3.41 ms, TE 2 = 5.87 ms, flip angle = 55^{o}, voxel size = 2.0 x 2.0 x 2.0 mm).

Anatomical reconstructions of each participant's brain were generated from the T1 structural image using Freesurfer v5.3 (Fischl, 2012). The functional data were first corrected for slice timing using the Fourier phase shift interpolation from SPM8 and then for B0 using FSL's *epidewarp* (http://www.nmr.mgh.harvard.edu/~greve/fbirn/b0/epidewarp.fsl). FS-FAST v5.3 was used for subsequent preprocessing with their default settings: coregistration with the

9

eNeuro Accepted Manuscript

corresponding Freesurfer anatomical reconstruction; motion correction to the first acquisition using the AFNI motion correction tool (http://afni.nimh.nih.gov/afni/); normalization to fsaverage/MNI space; and smoothing using 6 mm FWHM kernel.

fMRI Modeling and Analysis

Neuroimaging analyses were limited to *a priori* regions of interest in line with the literature (Talmi et al., 2009; Park et al., 2011; Bach et al., 2014; Aupperle et al., 2015; O'Neil et al., 2015; Schlund et al., 2016; Loh et al., 2017). Specifically, a cortical mask was constructed for left and right hemispheres using the Mindboggle atlas (Klein and Tourville, 2012) consisting of areas encompassing the cingulate cortex, dorsomedial prefrontal cortex, orbitofrontal cortex, dorsolateral and ventrolateral prefrontal cortex, and insular cortex (Figure 3). Similarly, a subcortical mask was constructed using the automated subcortical segmentation standard in Freesurfer (Fischl et al., 2002) consisting of the bilateral striatum (caudate, putamen), hippocampus, and amygdala.

In the first level analysis, we modeled the deliberation phase (time to response) using the variable epoch method (Grinband et al., 2008). The deliberation phase was modeled using two sets of boxcar regressors: one control regressor and one parametric modulation regressor. For both regressors, the boxcar for each trial was scaled in duration according to that trial's observed response time. The boxcar for each trial in the parametric modulation term was scaled in amplitude according to estimated decision conflict (d) for that trial. The parametric modulation boxcars corresponding to trials with missing responses were scaled to zero amplitude. Additionally, several separate control analyses were performed with the same procedure to determine the effect of 1) using the variable epochs method, 2) using an HBM to model individual differences and 3) using conflict as the parametric modulator over and above using

10

eNeuro Accepted Manuscript

risk or reward as the parametric modulator. For the first control analysis, fixed epochs were used instead of variable epochs, where the trial duration was not scaled and instead was uniform; from the presentation of the first stimulus (the risk bar) to 3.5 seconds after that time when subject responses were cut off. For the second control analysis, an equivalent non-hierarchical model was used (i.e. estimating only group parameters, excluding participant-level parameters) to model the conflict parametric modulator. For the third set of control analyses, risk and reward were used, in separate analyses, to parametrically modulate the deliberation-phase regressor instead of conflict, and, in another separate analysis, risk, reward and conflict were all used as parametric modulators with three parametrically modulated deliberation-phase regressors in the same first-level analysis. All regressors were convolved with the SPM hemodynamic response function. All estimated regression coefficients in first level analysis were converted to percent signal change (PSC; Pernet, 2014).

The fMRI data were preprocessed using a high-pass filter, nuisance regressors and motion scrubbing. A discrete cosine transform basis set was added to high-pass filter the data at 0.01 Hz. The six possible directions of motion were incorporated into the first-level analyses (after being demeaned, detrended, and orthogonalized) as nuisance regressors. Finally, motion scrubbing was used to mitigate the impact of high-motion acquisitions on the data (Siegel et al., 2014). Volumes for which the calculated framewise displacement (Power et al., 2012) exceeded 0.9 mm were excluded from analyses, and the first four acquisitions were discarded.

In the second level analysis, the beta coefficients estimated for each participant were submitted to a weighted least squares (WLS) regression where F-contrasts were computed for the control and parametrically modulated regressors. Scanner type (Trio vs. Prisma) was entered as a secondary nuisance regressor. Five thousand permutations of the WLS model were also

11

eNeuro Accepted Manuscript

computed following the Freedman–Lane procedure (Winkler et al., 2014). Every statistical map, both observed and permuted, was submitted to threshold-free cluster enhancement (Smith and Nichols, 2009) using the recommended parameters (H = 2, E = 0.5, step = 0.1). Finally, the permutation maps were used to compute family-wise error (FWE) corrections ($\alpha = 0.05$) for each voxel (Winkler et al., 2014). Any resulting clusters were discarded if they covered less than 100 mm^2 on the surface or fewer than 20 contiguous voxels of the volume.

Code Accessibility

All data and analysis scripts are available online at openneuro.org/datasets/ds001814 and github.com/mghneurotherapeutics/DARPA-ARC respectively. The data and scripts are freely available at these locations with instructions for access and suggested citation included.

Results

Behavioral Results

Participants exhibited the expected response trends for the ARC task: greater risk of electrical stimulation decreased on average the likelihood of selecting the risky option, whereas increasing reward increased the likelihood of selecting the risky option (Figure 3). The 95% highest density intervals (HDIs) of the posterior distribution for the group-level parameters showed decreases in risky-choice taking for the 50% ($\beta_1 = -1.922$, 95% HDI: [-2.606, -1.139]) and 90% risk ($\beta_2 = -4.180$, 95% HDI: [-5.273, -3.257]) conditions. In contrast, increases in risky-choice taking were observed in response to increasing reward ($\beta_3 = 10.652$, 95% HDI: [8.239, 12.887]).Thus, risk biased choice behavior towards avoidance (i.e. selecting the safe option), and reward biased choice behavior towards approach (i.e. selecting the risky option), indicating that the ARC task elicited the intended behavioral effects.

12

eNeuro Accepted Manuscript

At the subject level, the 95% HDIs of the posterior estimates for the 50% risk (β_1) and 90% risk (β_2) coefficients were strictly negative for 19/28 participants and 24/28 participants, respectively. The 95% HDIs for the reward coefficients (β_3) were strictly positive for 27/28 participants. No participants exhibited an increase in choice preference for the risky option with increasing risk and no participants exhibited an increase choice preference for the safe option with increasing reward. In summary, all the participants had response trends that matched our expectations for the ARC task, and most participants' behavior was modulated by both risk and reward.

For the response time component of our HBM, we found that approach-avoidance conflict was positively correlated with response times (Figure 4B). At the group-level, the 95% HDI of the posterior distribution on the conflict-RT slope parameter was strictly positive (α_1 = 0.456, 95% HDI: [0.388, 0.528]). The model estimated an average increase in response times of 0.456 seconds at maximal conflict. Thus, the ARC task was also successful in eliciting this hallmark behavioral signature of increased response times during approach-avoidance conflict.

It is important to note we observed considerable variability in the choice preferences of our participants (Figure 5). The most approach-biased participant selected the risky option on almost all trials (93%), whereas in contrast the most avoidance-biased participant selected the safe option on almost all trials (16%). This strongly demonstrates the notion that the points of maximal approach-avoidance conflict are unlikely to be the same across participants and reinforces the need for methods like HBMs that explicitly take into consideration these large individual differences.

Importantly, posterior predictive checks showed that our model accurately captured participants' choice behavior (Figure 5). The root-mean-square error between predicted and

13

eNeuro Accepted Manuscript

observed average risky choice was 0.023. To assess the possibility of model overfitting, we compared the widely applicable information criterion (WAIC; Watanabe, 2013) of our HBM to an equivalent non-hierarchical model (i.e. estimating only group parameters, excluding participant-level parameters). WAIC scores are reported here on deviance scale where lower scores denote greater fitness. The hierarchical model (WAIC = 1319.4) was strongly preferred to its non-hierarchical equivalent (WAIC = 2611.5) despite its greater complexity. We also compared our hierarchical model to a secondary hierarchical model that included interaction terms between risk and reward. This model performed slightly worse than the main effects-only model (WAIC = 1320.8). As such, we proceeded with the more parsimonious model with main effects only for fMRI analysis.

In summary, the ARC task successfully elicited approach, avoidance, and approach-avoidance conflict behaviors from all participants. Specifically, participants were (1) more likely to select the risky option with increasing reward; (2) more likely to select the safe option with increasing risk of electrical stimulation; and (3) slower to respond with increased approach-avoidance conflict. Moreover, participants exhibited large individual differences in their choice preferences, which were accurately captured by our HBM. It is worth reiterating that ignoring these differences can reduce contrast effects in fMRI analysis by averaging over the neural correlates of dissimilar cognitive processes (Ahn et al., 2011).

Imaging Results

For the control regressor (i.e. measuring the average BOLD signal change during the deliberation phase, without modulation by conflict), we found activations within the *a priori* cortical and subcortical regions of interest (Figure 6) that were selected based on prior literature (see the fMRI Modeling and Analysis section). Peak voxels and their corresponding statistics are

14

eNeuro Accepted Manuscript

reported in Table 1. Large, significant BOLD signal increase was observed in bilateral dorsal anterior cingulate and dorsomedial prefrontal cortex (dACC/dmPFC; BA 32), midcingulate cortex (BA 23/24), pre-supplementary motor area (pre-SMA; BA 6), anterior insula (BA 13), and dorsolateral prefrontal cortex (dlPFC; BA 46). Among subcortical structures, the control deliberation regressor was positively correlated with BOLD signal activation in bilateral dorsal hippocampus and striatum (caudate, putamen). Smaller, significant activations were also detected in the right lateral orbitofrontal cortex (OFC; BA 11) and right putamen. These results corroborate the distributed network of neural structures previously reported to be recruited during approach-avoidance conflict tasks (Talmi et al. 2009; Park et al., Bach et al., 2014; Aupperle et al., 2015; O'Neil et al., 2015; Schlund et al., 2016; Loh et al., 2017).

Significant change in BOLD signal for approach-avoidance conflict regressor was observed in a much more restricted set of structures (Figure 7,). Approach-avoidance conflict was positively correlated with BOLD signal activation only in bilateral rostral inferior frontal gyrus (IFG; pars orbitalis; BA 47), right dlPFC (BA 46), and right dmPFC/pre-SMA (BA 32). No significant positive activations were detected in subcortical structures, and no negative activations were detected in any *a priori* region of interest. In contrast to the aforementioned previous literature, our results suggest that only a select set of cortical structures tracked approach-avoidance conflict. Interestingly, our analysis revealed conflict representations in the right IFG, a structure previously unreported in the approach-avoidance conflict literature.

The control analyses showed the difference between these results and results from analyses with fixed epochs, averaging across subjects and using a simpler risk or reward only model. As shown in Figure 8, using fixed epochs caused smaller, more widespread, positive activations encompassing bilateral striatum and left insula in addition to the structures activated

15

eNeuro Accepted Manuscript

in the main, variable epochs analysis. The non-hierarchical Bayesian model (used in combination with variable epochs) had no significant activations that correlated with the conflict regressor. Using risk and reward as regressors in a model with only the non-parametrically modulated, control deliberation regressor and risk or the control deliberation regressor and reward also yielded almost no significant activations with the exception of a small, negative activation correlated with reward in a small area of right dlPFC and lateral OFC. When the risk and reward regressors were modeled in combination with conflict, not only were there no significant activations for the risk and reward regressors, but the significant activations for the conflict regressor was suppressed.

Discussion

In this study, we investigated the neural basis of human approach-avoidance conflict while accounting for two possible sources of heterogeneity in the literature; individual approach-avoidance variability and time-on-task. Using hierarchical Bayesian modeling, we controlled for individual differences in approach-avoidance preference by comparing participants' fMRI data according to each participant's relative points of maximal approach-avoidance conflict. Using the variable epochs method in our fMRI analyses, we also controlled for the time-on-task effect. Thus, we were able to differentiate brain structures strictly sensitive to approach-avoidance conflict from those representing information correlated with deliberation more generally. The present findings corroborate previous reports of the anatomical correlates of approach-avoidance behavior by our finding that BOLD signal increased during deliberation across a broad network of cortical and subcortical brain structures (dACC/dmPFC, pre-SMA, dlPFC, OFC, insula, striatum, hippocampus). Importantly, the current findings deviate from the previous literature insofar that our controlled analyses found conflict-related changes in BOLD signal only in a

16

eNeuro Accepted Manuscript

413 select set of structures (i.e. IFG, dlPFC, and pre-SMA). Collectively, the current findings suggest
414 the importance of careful methodology in isolating the neuroanatomical correlates of latent
415 psychological states such as approach-avoidance conflict.

416 To examine the effect of using an HBM, we compared these results to results from the
417 non-hierarchical Bayesian model analysis. The HMB methodology was clearly warranted by the
418 large differences observed between the approach-avoidance behavior of different participants as
419 shown in Figure 5 and described in the Behavioral Results section. The need for this
420 methodology was confirmed by the suppression of any significant areas of activation when a
421 non-hierarchical Bayesian model was used. Thus, accounting for individual differences with an
422 HBM resulted increased group-level fMRI contrast statistics, consistent with previous findings
423 (Ahn et al., 2011).

424 Another important difference between the present findings and past studies is our use of
425 the variable epochs method (Grinband et al., 2008), which we included so as to control for the
426 time-on-task effect and minimize the risk of mismodeling the hemodynamic response. By
427 controlling for time-on-task, our analysis was explicitly interested in identifying brain structures
428 that show an increase in the BOLD signal due to an increase in the intensity, not duration, of the
429 activity of the underlying neural populations. One natural question is whether approach-
430 avoidance conflict is more accurately modeled as the prolonged, but not increased, engagement
431 of brain structures. One problem with this view, as noted above, is that this makes it difficult to
432 disentangle conflict-specific signals from other correlated but unrelated signals (e.g. processing
433 of reward or threat stimuli). As such, we opted to use a more conservative definition of
434 approach-avoidance conflict (increase in amplitude of BOLD signal, above and beyond that
435 expected from prolonged engagement, as measured by our parametric modulation regressor).

17

eNeuro Accepted Manuscript

436 The conservativeness of this variable epochs method compared to fixed epochs was confirmed in
437 the control analyses shown in Figure 8, where areas with significant activation in the variable
438 epochs analysis were found to be a subset of areas with significant activation for the fixed epochs
439 analysis. Thus, our analysis was conservative but well suited to identify regions specifically
440 implicated in the processing of approach-avoidance conflict.

441 To control for whether our results relate to approach-avoidance and not some simpler
442 approach or avoidance alternative mechanism, we ran three different analyses 1) with risk as the
443 parametrically modulated regressor, 2) with reward as the parametrically modulated regressor
444 and 3) with three parametrically modulated regressors for risk, reward and conflict. The first two
445 analyses showed that risk or reward alone are not capable of explaining the regions of conflict
446 that had significant activations correlated with conflict (Figure 7); as described in the results
447 section, these analyses had almost no areas of significant activation. In the third analysis, the
448 suppression of significant conflict activations (described in the results section) suggested that
449 including risk and reward in the same model as conflict caused the variance to be split between
450 all three variables' explanatory power. Reward and risk are approach and avoidance stimuli,
451 respectively, so by definition these stimuli covary strongly with the approach-avoidance
452 measure—conflict. This control analysis therefore confirms that the explanatory power of
453 conflict is dependent on risk and reward and also shows that including regressors with high
454 covariance can cause a false-negative result.

455 Another point worth noting is that our analysis assumes only linear changes in the BOLD
456 response to conflict. The variable epochs method utilized here is insensitive to any nonlinear
457 changes in the BOLD signal that may arise as a function of response time, raising the possibility
458 of remaining biases in the present results. Interestingly, in a finite impulse response analysis of

18

eNeuro Accepted Manuscript

459 the hemodynamic response during prolonged response times, Yarkoni et al. (2009) found that
460 structures in the prefrontal cortex were better described by increases in the amplitude of
461 hemodynamic response but not by changes in its shape. These findings suggest that not using a
462 finite impulse response analysis did not bias the hemodynamic response in this present analysis,
463 but further studies are necessary to answer this question more definitively.

464 There were additional discrepancies between the present study and previous studies on
465 approach-avoidance tasks. Though positive BOLD activation was detected during deliberation in
466 the right orbitofrontal cortex (OFC), the effect was considerably smaller than previously reported
467 findings (Talmi et al., 2009; Schlund et al., 2016). This may reflect signal-to-noise ratio issues
468 particular to surface-based analysis of the OFC (Stenger, 2006). Additionally, in contrast to
469 Schlund et al. (2011) and Aupperle et al. (2015), amygdala activation was not detected during
470 deliberation. In this study, suboptimal calibration of the stimulation amperage likely diminished
471 participants' perception of threat from the stimulation and consequently their amygdala
472 activation. Finally, the bilateral hippocampus activations detected during deliberation were
473 located dorsally, rather than anteriorly/ventrally as have been previously reported in literature on
474 threat processing (Bach et al., 2014). The dorsal hippocampus has been associated with cognition
475 and planning (Fanselow and Dong, 2010), so these activations could reflect participants'
476 processing of the conditional structure of the ARC task (e.g. "if *safe* is chosen, then 0% chance
477 of electrical stimulation; if *risky* is chosen, then X% chance of electrical stimulation").

478 This study had several limitations. Due to the equipment issues described above, as well
479 as the use of non-adaptive rewards, we were unable to calibrate the reward and risk of the ARC
480 task according to each participant's choice preferences. This may be one reason why we
481 observed an approach-bias on average. This also means that the present study undersampled

19

eNeuro Accepted Manuscript

trials at or near the points of participants' maximal approach-avoidance conflict. A consequence of this undersampling is that many of the high conflict decisions participants made in this task occurred during high risk trials, making it harder to divorce conflict from risk. Future approach-avoidance conflict experiments should consider incorporating adaptive design optimization (Myung et al., 2013) in order to titrate the levels of rewarding and threatening stimuli according to future participants' choices preferences to minimize the influence of these potential biases.

Finally, it is worth noting that the set of structures we found correlated with approach-avoidance conflict (i.e. IFG, dlPFC and pre-SMA) share overlap with the putative response inhibition network (Aron et al., 2004, 2014; Aron and Poldrack, 2006). One interpretation of the present results is that approach-avoidance conflict is another process requiring response inhibition, wherein the IFG inhibits prepotent motor responses in order to facilitate prolonged evidence accumulation during difficult choices. This interpretation is consistent with the increased response times observed in the present experiment. The possible role of the inhibition network during approach-avoidance conflict points to a clear direction for future studies; investigating whether the putative response inhibition network works to signal response conflict to other brain structures, such as through the hyperdirect pathway to the basal ganglia (Frank et al., 2015). Alternately, these structures may be involved in the resolution of approach-avoidance conflict, such as by biasing choice towards approach or avoidance. In either case, the framework that this study presents for the consideration of individual-level behavioral variation and the time-on-task effect would likely lead to benefits in specificity and accuracy of future studies investigating similar cognitive processes.

20

eNeuro Accepted Manuscript

References

Ahn W-Y, Krawitz A, Kim W, Busmeyer JR, Brown JW (2011) A Model-Based fMRI Analysis with Hierarchical Bayesian Parameter Estimation. J Neurosci Psychol Econ 4:95–110.

Amemori K-I, 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.

American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Aron AR, Poldrack RA (2006) Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. J Neurosci 26:2424–2433.

Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8:170–177.

Aron AR, Robbins TW, Poldrack RA (2014) Inhibition and the right inferior frontal cortex: one decade on. Trends Cogn Sci 18:177–185.

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.

Aupperle RL, Sullivan S, Melrose AJ, Paulus MP, Stein MB (2011) A reverse translational approach to quantify approach-avoidance conflict in humans. Behav Brain Res 225:455–463.

Bach DR, Guitart-Masip M, Packard PA, Miró J, Falip M, Fuentemilla L, Dolan RJ (2014) Human hippocampus arbitrates approach-avoidance conflict. Curr Biol 24:541–547.

Badre D (2008/5) Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. Trends Cogn Sci 12:193–200.

Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017) Stan: A Probabilistic Programming Language. J Stat Softw 76:1–32.

Carver CS, White TL (1994) Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. J Pers Soc Psychol 67:319.

Corr PJ (2013) Approach and Avoidance Behaviour: Multiple Systems and their Interactions. Emot Rev 5:285–290.

Dale AM, Buckner RL (1997) Selective averaging of rapidly presented individual trials using fMRI. Hum Brain Mapp 5:329–340.

Dixon ML, Fox KCR, Christoff K (2014) Evidence for rostro-caudal functional organization in multiple brain areas related to goal-directed behavior. Brain Res 1572:26–39.

Elliot AJ (2008) Handbook of Approach and Avoidance Motivation. Taylor & Francis.

Fanselow MS, Dong H-W (2010) Are the dorsal and ventral hippocampus functionally distinct structures? Neuron 65:7–19.

Fischl B (2012) FreeSurfer. NeuroImage 62:774–781.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33:341–355.

Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, Cavanagh JF, Badre D (2015) fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. J Neurosci 35:485–494.

Friedman A, Homma D, Gibb LG, Amemori K-I, Rubin SJ, Hood AS, Riad MH, Graybiel AM (2015) A Corticostriatal Path Targeting Striosomes Controls Decision-Making under Conflict. Cell 161:1320–1333.

Gelman A, Carlin JB, Stern HS, Rubin DB (2014) Bayesian data analysis. Chapman & Hall/CRC Boca Raton, FL, USA.

Gelman A, Jakulin A, Pittau MG, Su Y-S (2008) A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat 2:1360–1383.

Ghosh J, Li Y, Mitra R (2017) On the Use of Cauchy Prior Distributions for Bayesian Logistic Regression. Bayesian Anal Available at: http://projecteuclid.org/euclid.ba/1488855634 [Accessed June 4, 2017].

Gramfort, A, Luessi M, Larson E, Engemann D, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen M (2014) MNE software for preprocessing MEG and EEG data, NeuroImage, Volume 86, 1 February 2014, Pages 446-460, ISSN 1053-8119.

Gramfort A, Luessi M, Larson E, Engemann D, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python, Frontiers in Neuroscience, Volume 7, 2013, ISSN 1662-453X.

Grinband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J (2008) Detection of time-varying signals in event-related fMRI designs. Neuroimage 43:509–520.

21

eNeuro Accepted Manuscript

Klein A, Tourville J (2012) 101 labeled brain images and a consistent human cortical labeling protocol. Front Neurosci 6:171.

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C, Others (2007) What's new in Psychtoolbox-3. Perception 36:1.

Kruschke JK (2015) Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. Academic Press.

Loh E, Kurth-Nelson Z, Berron D, Dayan P, Duzel E, Dolan R, Guitart-Masip M (2017). Parsing the Role of the Hippocampus in Approach-Avoidance Conflict. *Cerebral Cortex*, Vol 27, 1:201–215.

Myung JI, Cavagnaro DR, Pitt MA (2013) A Tutorial on Adaptive Design Optimization. J Math Psychol 57:53–67.

O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci 1104:35–53.

O'Neil EB, Newsome RN, Li IHN, Thavabalasingam S, Ito R, Lee ACH (2015) Examining the Role of the Human Hippocampus in Approach-Avoidance Decision Making Using a Novel Conflict Paradigm and Multivariate Functional Magnetic Resonance Imaging. JNeurosci 35 (45) 15039-15049.

Park SQ, Kahnt T, Rieskamp J, Heekeren HR. Neurobiology of value integration: when value impacts valuation. J Neurosci. 2011;31:9307–9314.

Pedersen ML, Frank MJ, Biele G (2016) The drift diffusion model as the choice rule in reinforcement learning. Psychon Bull Rev Available at: http://dx.doi.org/10.3758/s13423-016-1199-y.

Pernet CR (2014) Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers. Front Neurosci 8:1.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–2154.

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.

Schlund MW, Magee S, Hudgins CD (2011) Human avoidance and approach learning: evidence for overlapping neural systems and experiential avoidance modulation of avoidance neurocircuitry. Behav Brain Res 225:437–448.

Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R, Dunbar GC (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 59 Suppl 20:22–33;quiz 34–57.

Siegel JS, Power JD, Dubis JW, Vogel AC, Church JA, Schlaggar BL, Petersen SE (2014) Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. Hum Brain Mapp 35:1981–1996.

Sierra-Mercado D, Deckersbach T, Arulpragasam AR, Chou T, Rodman AM, Duffy A, McDonald EJ, Eckhardt CA, Corse AK, Kaur N, Eskandar EN, Dougherty DD (2014) Decision making in avoidance–reward conflict: a paradigm for non-human primates and humans. Brain Struct Funct 220:2509–2517.

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83–98.

Stan Development Team (2017) Prior Choice Recommendations. Available at: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations [Accessed June 4, 2017].

Stenger VA (2006) Technical considerations for BOLD fMRI of the orbitofrontal cortex. The Orbitofrontal Cortex Oxford University Press, Oxford:423–446.

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.

Taylor JSH, Rastle K, Davis MH (2014) Interpreting response time effects in functional imaging studies. Neuroimage 99:419–433.

van der Kouwe AJW, Benner T, Salat DH, Fischl B (2008) Brain morphometry with multiecho MPRAGE. Neuroimage 40:559–569.

Vehtari A, Gelman A, Gabry J (2016) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput:1–20.

Watanabe S (2013) A Widely Applicable Bayesian Information Criterion. J Mach Learn Res 14:867–897.

Widge AS, Ellard KK, Paulk AC, Basu I, Yousefi A, Zorowitz S, Gilmour A, Afzal A, Deckersbach T, Cash SS, Kramer MA, Eden UT, Dougherty DD, Eskandar EN (2017) Treating refractory mental illness with closed-loop brain stimulation: Progress towards a patient-specific transdiagnostic approach. Exp Neurol Available at: http://dx.doi.org/10.1016/j.expneurol.2016.07.021.

22

eNeuro Accepted Manuscript

Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014) Permutation inference for the general linear model. Neuroimage 92:381–397.

Yarkoni T, Barch DM, Gray JR, Conturo TE, Braver TS (2009) BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis. PLoS One 4:e4257.

Yousefi A, Paulk AC, Deckersbach T, Dougherty DD, Eskandar EN, Widge AS, Eden UT (2015) Cognitive state prediction using an EM algorithm applied to Gamma distributed data. Conf Proc IEEE Eng Med Biol Soc 2015:7819–7824.

Zeidan MA, Igoe SA, Linnman C, Vitalo A, Levine JB, Klibanski A, Goldstein JM, Milad MR (2011) Estradiol modulates medial prefrontal cortex and amygdala activity during fear extinction in women and female rats. Biol Psychiatry 70:920–927.

23

eNeuro Accepted Manuscript

Legends

Figure 1: The Aversion-Reward Conflict (ARC) task. Participants are presented with a safe choice (blue) and a risky choice (orange). The safe choice pays a guaranteed small reward ($0.01) and no aversive stimulation. The risky choice pays a guaranteed larger reward ($0.05 - $0.95), and a probability of stimulation as indicated by the centered white bar. Participants decide whether to accept a higher payout at risk of aversive stimulation.
Figure Contributions: Darin Dougherty, Thilo Deckersbach, Alik Widge and Samuel Zorowitz designed the task, Sam Zorowitz created the figure.

Figure 2: A Kruschke-style diagram of the hierarchical model. The "~" symbol indicates stochastic dependency, whereas the "=" symbol indicates a deterministic dependency. Ellipses indicate the indices over which the dependency applies. The parameter of most interest is *d*, the inverse distance-to-decision-boundary, which measures the estimated conflict experienced on a given trial.
Figure Contributions: Samuel Zorowitz created the model.

Figure 3: *A priori* cortical regions of interest. Regions (Freesurfer labels) were selected from the Mindboggle atlas (mindboggle.info/data.html) based on the diffuse locations of activations previously reported in the approach-avoidance decision making literature.
Figure Contributions: Samuel Zorowitz chose the regions of interest based on prior literature and created the figure.

Figure 4: Group-level behavior results. (A) The estimated likelihood of choosing the risky option for each risk level and across rewards. The model estimated decreases in risky decision making at both 50% risk (β_1 = -1.922, 95% HDI: [-2.606, -1.139]) and 90% risk (β_2 = -4.180, 95% HDI: [-5.273, -3.257]). In contrast, the model estimated increases in risky decision making in response to increasing reward (β_3 = 10.652, 95% HDI: [8.239, 12.887]). (B) The estimated linear component of deliberation time as a function of decision conflict, *d*. The model estimated an increase in deliberation time with decision conflict (α_1 = 0.456, 95% HDI: [0.388, 0.528]). Shaded regions denote the 95% highest density interval (HDI).
Figure Contributions: Samuel Zorowitz, Katherine Link and Alexander Rockhill performed the behavioral analysis.

Figure 5: Individual differences in behavior. Participants in the ARC task exhibited large individual differences in behavior. (A) Participants varied in their approach-avoidance preferences (though the majority were approach-biased). (B) Participants varied in the extent to which their deliberation increased in response to decision conflict (but all participants showed increased response times during conflict). Each point represents one participant. The horizontal axis denotes the observed behavior (proportion of risky choices, A; response time increases, B), and the vertical axis denotes the model predicted behavior. Proximity to the diagonal indicates goodness of fit.
Figure Contributions: Samuel Zorowitz, Katherine Link and Alexander Rockhill performed the behavioral analysis.

24

eNeuro Accepted Manuscript

Figure 6: Percent signal change (PSC) during deliberation. The control regressor measures changes in the BOLD signal during deliberation (independent of approach-avoidance conflict). Positive activation was found in cortical and subcortical regions including the lateral and medial prefrontal cortex, striatum, and hippocampus. All voxels corrected for multiple comparisons through 5000-iteration permutation testing and voxel-wise FWE corrections ($\alpha = 0.05$).
Abbreviations: LH = left hemisphere; RH = right hemisphere.
Figure Contributions: Samuel Zorowitz and Alexander Rockhill performed the fMRI analysis; Samuel Zorowitz, Alexander Rockhill and Kristen Kellard collected the data.

Figure 7: Percent signal change (PSC) during conflict. The parametric modulation regressor measures changes in BOLD signal during deliberation as a function of approach-avoidance conflict. Positive activation was detected only in bilateral inferior frontal gyrus (IFG), and right dorsolateral prefrontal cortex (dlPFC) and pre-supplementary motor area (pre-SMA). All voxels corrected for multiple comparisons through 5000-iteration permutation testing and voxel-wise FWE corrections ($\alpha = 0.05$).
Figure Contributions: Samuel Zorowitz and Alexander Rockhill performed the fMRI analysis; Samuel Zorowitz, Alexander Rockhill and Kristen Kellard collected the data.

Figure 8: Percent signal change (PSC) during conflict for the fixed epochs analysis. In this case, epochs were made from the first stimulus presentation to the end of the response period instead ending when the subject responded for each particular trial. More widespread, less specific, smaller, positive activation was detected in the same structures as Figure 7 with the addition of activation in bilateral striatum, left insula as well as greater activation in bilateral dlPFC. All voxels corrected for multiple comparisons through 5000-iteration permutation testing and voxel-wise FWE corrections ($\alpha = 0.05$).
Figure Contributions: Alexander Rockhill performed the fMRI analysis; Samuel Zorowitz, Alexander Rockhill and Kristen Kellard collected the data.

Table 1: Coordinates and statistics of peak BOLD activations. The reported statistics are the percent signal change (PSC) and weighted-least squares (WLS) contrast against baseline (*F*) statistic. The first set of results reflect the unmodulated deliberation and the second set reflect the contrast between deliberation parametrically modulated by conflict and unmodulated deliberation. All coordinates reported in the Montreal Neurological Institute (MNI) space and reflect the peak of activation. All voxel statistics were corrected for multiple comparisons through 5000-iteration permutation testing and voxel-wise FWE corrections ($\alpha = 0.05$).
Abbreviations: LH = left hemisphere; RH = right hemisphere; dACC = dorsal anterior cingulate cortex; dmPFC = dorsomedial prefrontal cortex; MCC = midcingulate cortex; DLPFC = dorsolateral prefrontal cortex; OFC = orbitofrontal cortex; pre-SMA = pre-supplementary motor cortex; IFG = inferior frontal gyrus.
Figure Contributions: Samuel Zorowitz and Alexander Rockhill performed the fMRI analysis; Samuel Zorowitz, Alexander Rockhill and Kristen Kellard collected the data.

25

eNeuro Accepted Manuscript

709 Tables
710

Deliberation Phase (Control)					
ROI	**X**	**Y**	**Z**	**PSC**	**F**
dACC/dmPFC: LH	-12	22	36	0.08	352.92
RH	7	15	24	0.09	462.05
MCC: LH	-7	-22	29	0.15	328.72
RH	7	-15	31	0.18	529.27
pre-SMA: LH	-9	7	51	0.10	419.56
RH	10	14	47	0.10	373.34
dlPFC: LH	-36	9	24	0.12	223.96
RH	36	18	25	0.11	312.87
Anterior Insula: LH	-31	27	9	0.2	351.91
RH	31	27	8	0.16	413.00
Lateral OFC: RH	13	38	-24	0.07	95.60
pre-Motor: LH	-37	-2	43	0.14	291.16
RH	36	-3	44	0.14	333.73
Caudate: LH	-10	7	3	0.07	28.42
RH	10	11	5	0.06	25.16
Putamen: LH	-20	5	1	0.05	29.04
RH	34	-7	-7	0.04	22.15
Hippocampus: LH	-14	-39	-3	0.09	34.95
RH	14	-39	-1	0.10	34.47

26

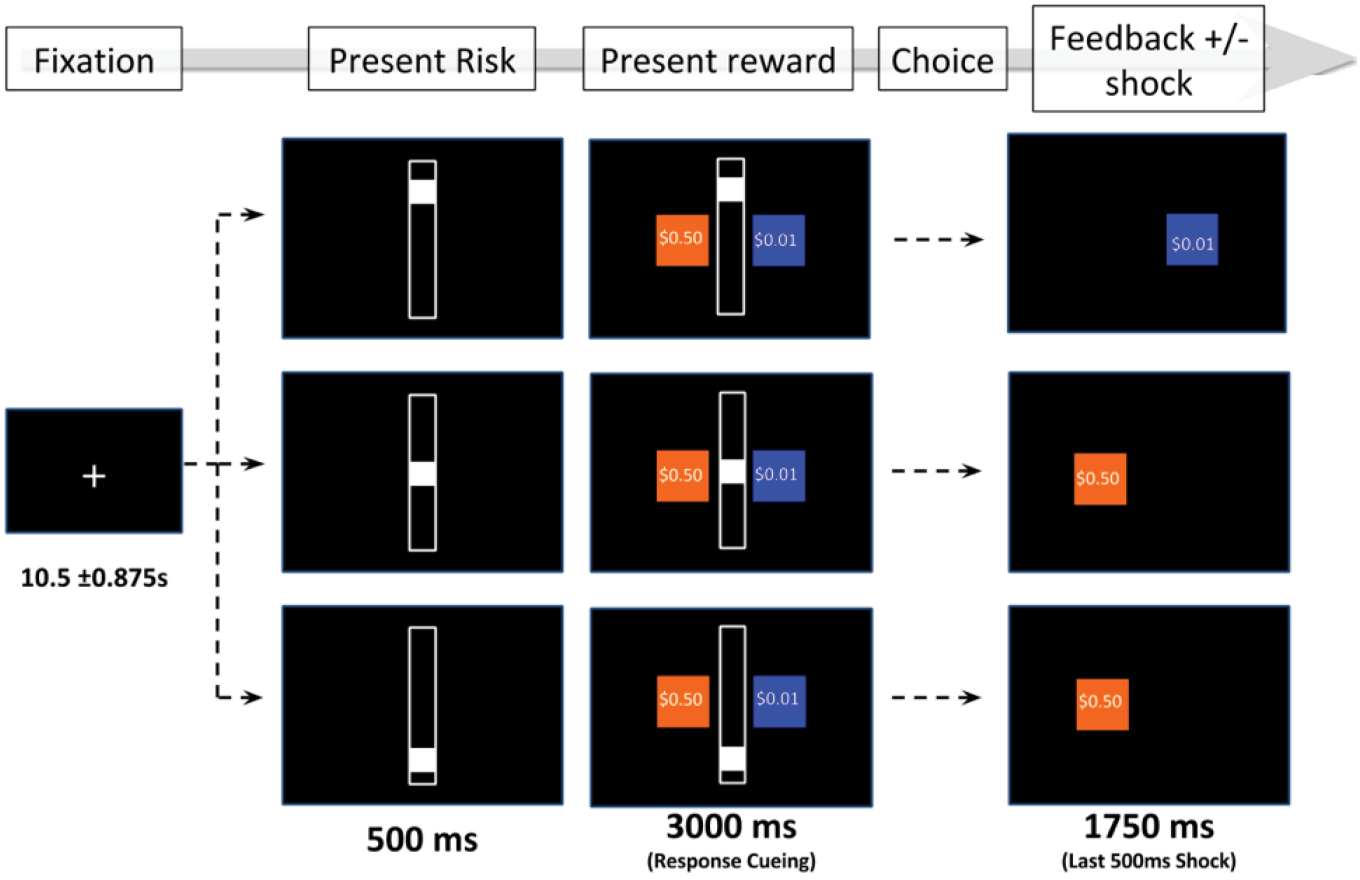
eNeuro Accepted Manuscript

Deliberation Phase (Conflict)					
IFG: LH	-39	45	7	0.05	56.53
RH	42	45	-6	0.05	55.10
dlPFC: RH	42	27	31	0.04	68.02
pre-SMA: RH	9	27	46	0.04	59.55

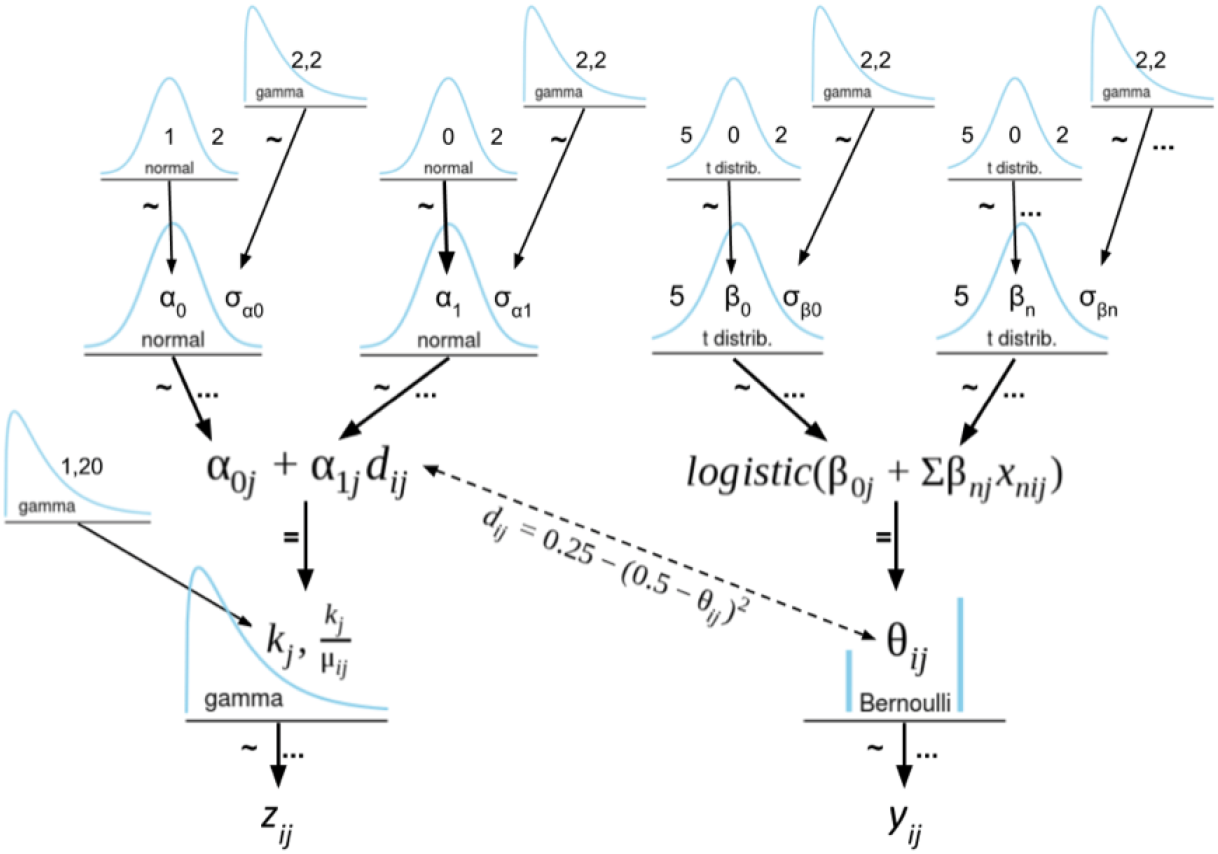
711

27

eNeuro Accepted Manuscript

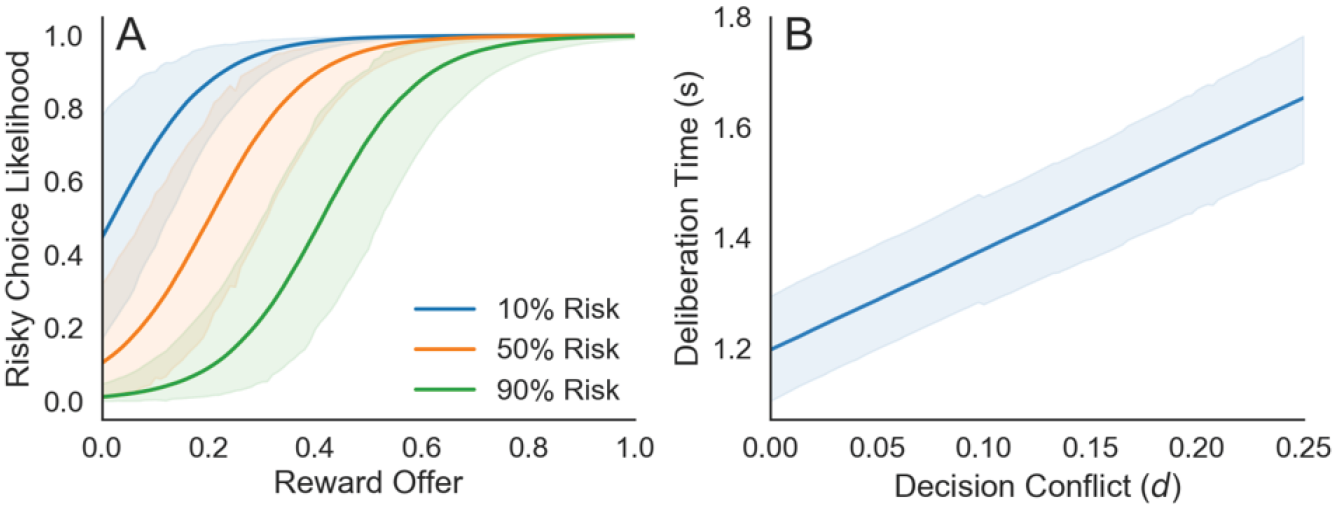


eNeuro Accepted Manuscript

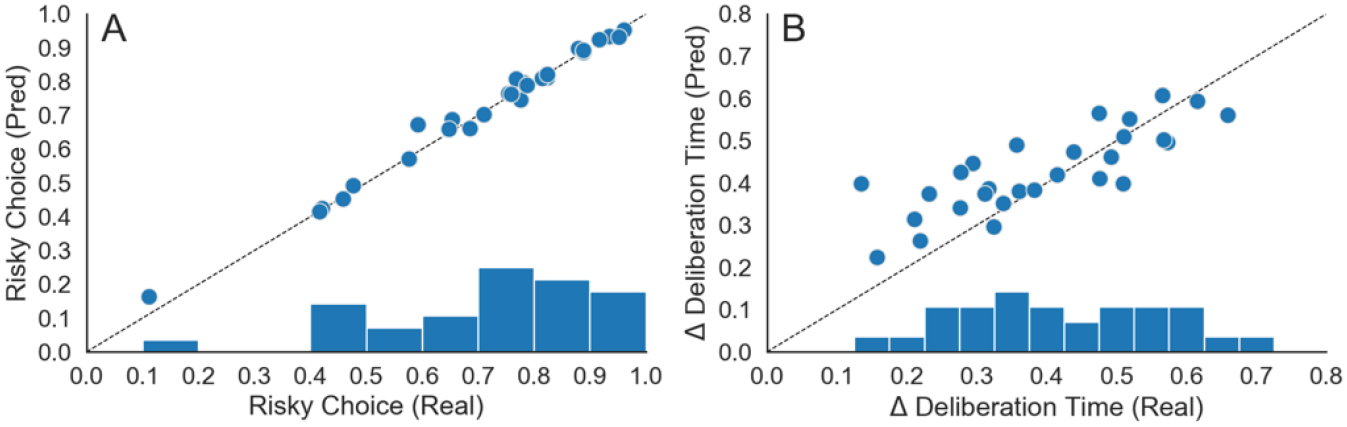


eNeuro Accepted Manuscript

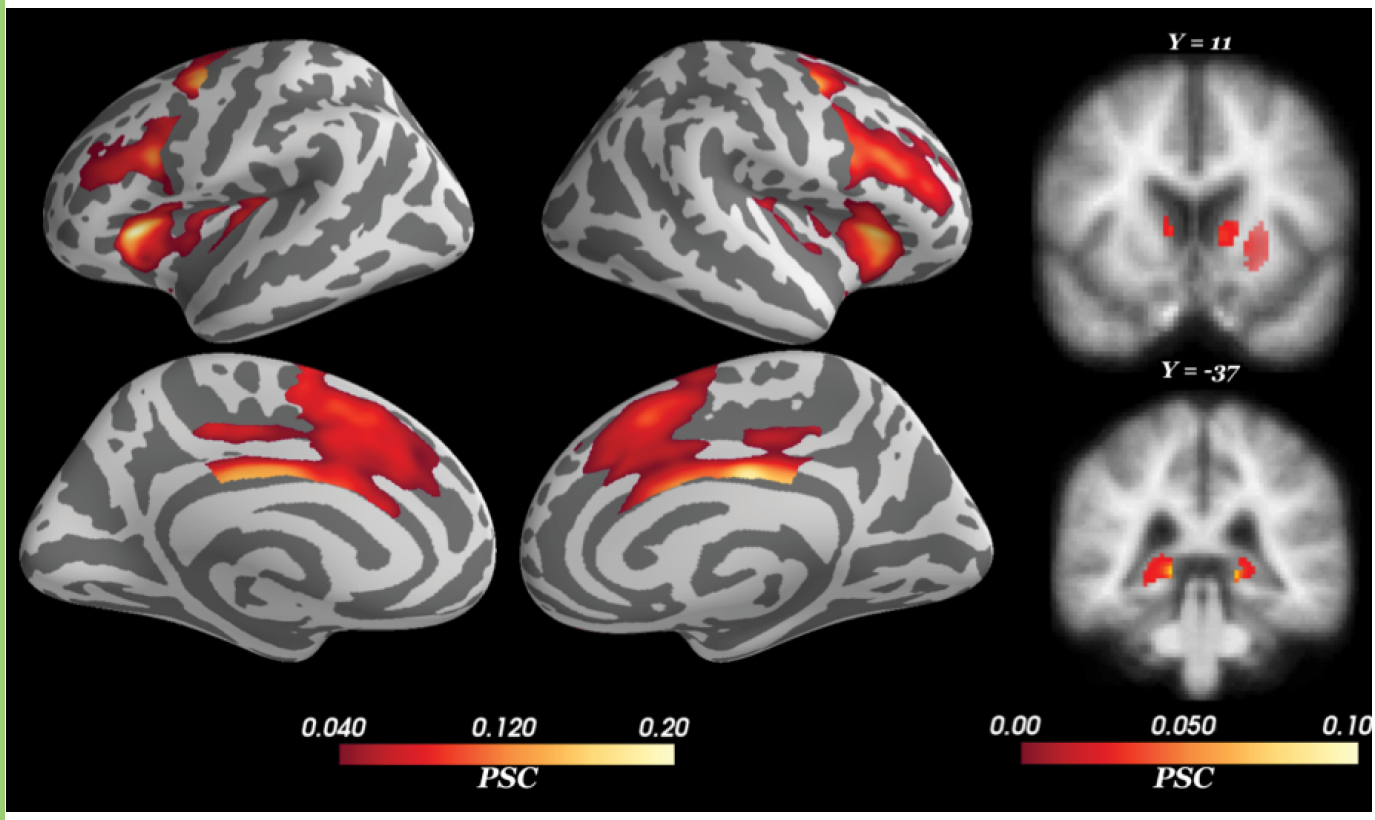
eNeuro Accepted Manuscript



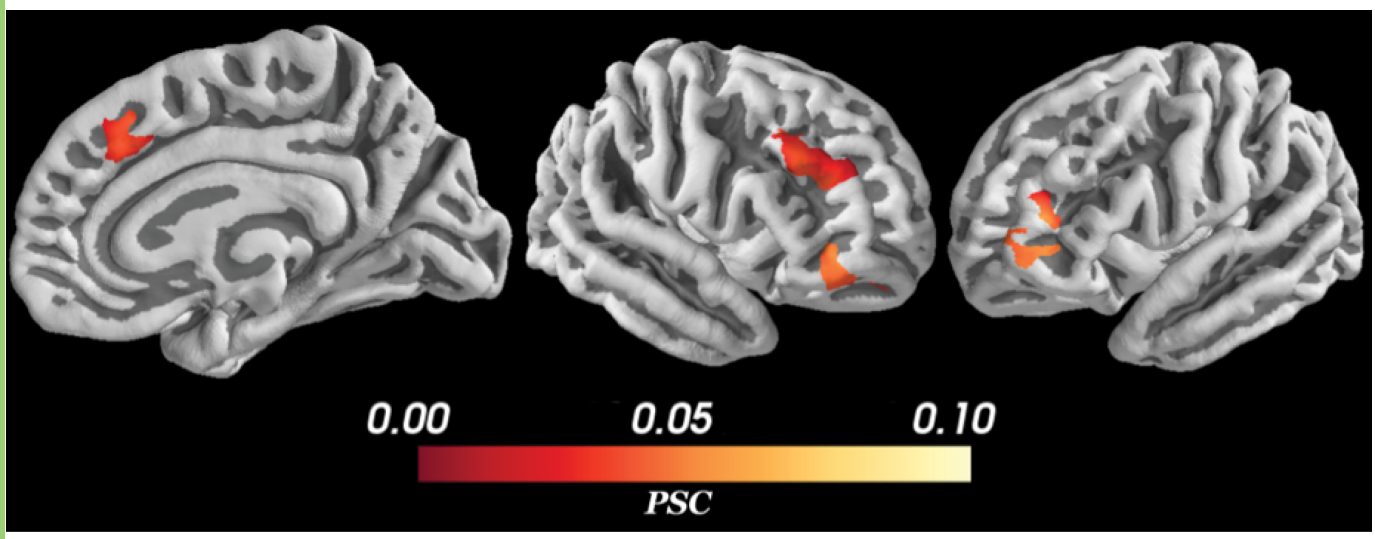
eNeuro Accepted Manuscript



eNeuro Accepted Manuscript



eNeuro Accepted Manuscript



eNeuro Accepted Manuscript