Brain Network Connectivity Underlying Decisions Between the "Lesser of Two Evils"

Colleen Mills-Finnerty, Ph.D.*, Catherine Hanson, Ph.D.**, Stephen Jose Hanson, Ph.D.**
*Stanford University Dept. of Psychiatry and Behavioral Science
**Rutgers University Newark Dept. of Psychology

Keywords: aversive, framing, connectivity, prefrontal cortex, striatum

Corresponding author information:

Colleen Mills-Finnerty, Ph.D.

Stanford University Dept. of Psychiatry and Behavioral Science
401 Quarry Road

Stanford, CA

cmfinn@stanford.edu

Abstract

In daily life we are often forced to choose between the "lesser of two evils," yet there remains limited understanding of how the brain encodes choices between aversive stimuli, particularly choices involving hypothetical futures. We tested how choice framing affects brain activity and network connectivity by having participants make choices about individualized, aversive, hypothetical stimuli (i.e. illnesses, car accidents, etc.) under approach and avoidance frames ("which would you rather have/avoid") during fMRI scanning. We tested whether limbic and frontal regions show patterns of signal intensity and network connectivity that differed by frame, and compared this to response to similar appetitive choices involving appetitive preferences (i.e. hobbies, vacation destinations). We predicted that regions such as the insula, amgydala, and striatum would respond differently to approach vs. avoidance choices during aversive hypothetical choices. We identified activations for both choice frames in areas broadly associated with decision making, including the putamen, insula, and anterior cingulate, as well as deactivations in areas shown to be sensitive to valence, including the amygdala, insula, prefrontal cortex, and hippocampus. Connectivity between brain regions differed based on choice frame, with greater connectivity among deactive regions including the amygdala, insula, and ventromedial prefrontal cortex during avoidance frames compared to approach frames. These differences suggest that approach and avoidance frames lead to different behavioral and brain network response when deciding which of two evils are the lesser.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

Introduction

Unpleasant decisions are part of everyday life, whether it's choosing which bill to pay first, which painful medical treatment to pursue, or perhaps which candidate to vote for. Often these choices involve hypothetical future outcomes, such as potential recovery time from surgery. The biases that may influence choices between the "lesser of two evils" are important to characterize to understand how and why people make choices that may seem to be against their best interests or violate maxims of rationality. Here we aim to bridge between behavioral economic models of choice and real-world decision making behavior by studying a well established choice bias, the framing effect, in the context of decisions about real-world relevant aversive stimulus categories such as illnesses and car accidents. We characterize behavior, BOLD magnitude, as well as connectivity relationships implicated in the processing of such choices.

It is well-established that when choices are presented with emphasis on potential loss, people make different decisions than when the same choices are presented in terms of potential gains or positive outcomes (i.e. framing effects; Tversky & Kahneman, 1986). However, while these framing effects are well characterized in domains such as financial rewards and losses, it is less clear how such framing effects impact hypothetical aversive choices. An additional issue is that stimulus valence may impact how the brain encodes choice options, with some areas processing primarily salience or intensity by increasing activation for both highly appetitive and aversive stimuli, whereas other areas may demonstrate valence sensitivity by increasing magnitude for appetitive stimuli and decreasing it for negative. Here, we characterize brain dynamics underlying framing effects on complex, real world relevant, hypothetical aversive choices in terms of both magnitude based dynamics and magnitude-independent connectivity to

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

test whether areas implicated in decision making respond to salience, valence, or both.

While brain response to actual aversive stimuli is relatively well described, dynamics underlying hypothetical aversive choices are less so. For example, "real" stimuli used in previous studies include unappealing foods or beverages (Harris et al., 2011; Kang & Camerer, 2013; Metereau et al., 2014), negative feedback (Bhanji & Delgado, 2014), electrical shocks (Collins et al., 2014; Lawson et al., 2014; Winston et al., 2014) monetary losses (e.g. Delgado et al., 2003; Kahnt et al., 2014), tactile stimulation (e.g. uncomfortable heat, pressure, or textures; Roy et al., 2014; Lamm et al., 2014), odors (Gottfried et al., 2002), and unattractive faces (Martín-Loeches et al, 2014). However, fewer studies have addressed whether brain response when actually experiencing an aversive stimulus is different from choosing amongst hypothetical aversive stimuli (e.g. Sharot et al., 2010; Feldmen-Hall et al., 2012; Kang & Camerer, 2013), measured response to multiple types of aversive stimuli in the same subjects (e.g. Lamm et al., 2015; Metereau et al., 2014), or attempted to simulate real-world aversive choice scenarios in the lab (e.g. Sharot et al., 2010). This distinction is important because the process of dealing with an actual negative outcome in the "here and now" may differ from making choices about the same outcome in the hypothetical future (Benoit et al., 2014, Gerlach et al, 2011), and real versus hypothetical choices can involve recruitment of different brain networks, for example hypothetical moral choices may rely more heavily on an "imagination" network than "real" choices that result in an immediate outcome (Feldman Hall et al., 2012). Previous research has established that real versus hypothetical choices for appetitive stimuli recruit similar brain networks (Mills-Finnerty et al., 2014) but it not clear whether this is the case for hypothetical aversive stimuli. According to several recent meta-analyses, areas that may be specialized for processing the value of actual aversive stimuli include posterior cingulate, amygdala,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

parahippocampus, and inferior frontal gyrus; areas selective for appetitive stimuli may include anterior cingulate and superior temporal gyrus; and areas that may play a role in both include thalamus, amygdala, hippocampus, insula, ventral striatum, and certain regions of ventromedial prefrontal cortex (VMPFC) and dorsolateral prefrontal cortex (DLPFC; Liu et al., 2011; Hayes et al., 2014; Lindquist et. al, 2014). Several of these areas also play well established roles in conflict-based decision making more generally, such as the striatum and anterior cingulate (Botvinick, 2007; Brown & Alexander, 2013; Kolling et al., 2014; Friedman et al., 2015; Robertson et al., 2015). Here we aim to clarify whether hypothetical aversive choices recruit similar brain areas as actual aversive choices by adapting the choice paradigm from Mills-Finnerty et al. (2014) to involve choices for hypothetical aversive stimuli. Brain response during aversive hypothetical choice is then compared against that for hypothetical appetitive choice to clarify whether 1. the same network of regions is broadly involved; 2. if those areas demonstrate involvement via activation increases, decreases or both; and 3. if and how network connectivity shifts in response to differences in choice frame and stimulus valence.

Choice framing can influence decisions such that choices where the emphasis is placed on gain elicit different responses than choices where the emphasis is placed on loss (Kahneman & Tversky, 1981). Framing effects have been well studied in terms of both behavior (see Kuhberger, 1998 for meta-analysis) and brain response, with evidence of involvement of the amygdala (DeMartino et al., 2006), striatum and ventromedial prefrontal cortex (Tom et al., 2007; Foo et al., 2014), and dorsolateral prefrontal cortex (Foo et al., 2014). Loss frames tend to encourage riskier decisions than gain frames, due to loss aversion, whereby offsetting a loss requires a gain twice as large. Under the threat of loss, riskier decisions may become more appealing if they offer the chance at avoiding a loss altogether. The amygdala has been

implicated in loss aversion, with both lesion patients and rats with amygdala lesions showing diminished loss aversion response (DeMartino et al., 2010; Tremblay et al., 2014) and evidence that loss magnitude is tracked via signal in the amygdala and insula (Canessa et al., 2013). Additionally, regions associated with decision making such as DLPFC, VMPFC, anterior cingulate cortex (ACC), insula and striatum shift their response magnitude and connectivity patterns based on whether a choice is framed as positive/gain based or negative/loss based (Foo et al., 2014; Mills-Finnerty et al., 2014). For example, in one study using monetary gambles, increased activation in orbital and medial prefrontal cortex was correlated with decreased susceptibility to framing, meaning less bias towards risky decisions during loss frames (DeMartino et al., 2006). Participants making judgements about self relevant descriptors such as cleverness or honesty were more likely to endorse positively framed statements (i.e. “I am honest at least 75% of the time”) than negative (“I am not honest up to 25% of the time”). Positively framed judgements were related to greater mPFC activation, whereas negative judgements activated regions such as the insula (Murch & Krawczyk, 2014). Previous work has used framing manipulations with hypothetical, high complexity appetitive stimuli (Mills-Finnerty et al., 2014) or appetitive and aversive foods (Foo et al., 2014) but no studies have compared framing effects on appetitive and aversive multidimensional and hypothetical stimuli using connectivity modelling. A key question is whether avoiding a hypothetical negative stimulus (negative “reinforcement”) involves similar mechanisms in terms of magnitude and connectivity as approaching a positive stimulus (positive “reinforcement”). Here, we test how framing scenarios as approaching or avoiding hypothetical aversive stimuli affects behavior and brain response to clarify dynamics underlying these processes. Specifically, we test whether avoiding a hypothetical negative outcome recruits the same brain regions (e.g. striatum, mPFC) as

approaching an appetitive hypothetical or real reward, by comparing magnitude based changes during appetitive vs. aversive choices and examining connectivity-based changes in response to frame in the aversive domain.

Replicating the complexity of real world aversive scenarios is challenging to do in an experimentally robust way. Common frameworks such as using food or money rewards offer simple and standardized scaling of stimulus dimensions (e.g. monetary value, calories) but therefore do not capture the multidimensional nature of naturalistic choices. Here we use a novel, multidimensional, individualized stimulus set to better approximate the complexity of real world decision making. This also enables the use of mixed effects modeling and generalizable results, as opposed to most task stimuli which are more appropriately modeled as fixed effects (Westfall et al, 2016). Since in our task all aversive choices are hypothetical, we are not limited to using stimuli like shocks or odors and so instead ask participants about scenarios such as contracting types of illnesses or experiencing types of car accidents. Unlike stimuli such as electric shocks or monetary losses, hypothetical choices avoid the confound of hedonic/sensory elements of pain, the logistical issues of implementing actual losses in the lab (such as monetary penalties), and the artificiality of using stimuli such as shocks. Disentangling the valence of stimuli from the outcome they predict (since no outcomes are expected or actually occur during our task) also removes potentially confounding explicit goal motivations. Since these are hypothetical scenarios where choice behavior does not lead an outcome, participant's choices can instead be used to infer preferences in a revealed preference framework; e.g. things chosen to be avoided all the time are interpreted as being preferred less than things only avoided sometimes. Therefore choices here are interpreted as the behavioral readout of a process we believe reflects preferences, or judgements, such as "X is worse/better than Y." This approach allows us to

customize stimuli to participant's *perception* of severity through the use of individualized stimulus categories. We refer to these hypothetical, multi dimensional, individualized aversive stimuli as "abstract reinforcers" to distinguish them from concrete reinforcers such as immediate delivery of money, food, or shocks, reinforcement here referring to the internal positive or negative processing that may motivate approach or avoidance behavior (e.g. relief from escaping a negative outcome).

We make several predictions about the general effects of stimulus valence on choice: we expect that consistent with response in the appetitive domain (Mills-Finnerty et al., 2014), changes in activation will be observed in brain regions associated with decision making such as the striatum, mPFC, insula, and amygdala, during choices for hypothetical aversive stimuli. Behaviorally, we expect that avoidance frames will result in faster decision times than approach frames, under the assumption that it is easier to decide which aversive stimulus to avoid than approach, an account consistent with previous literature (e.g. Kim et al. 2006, Fitzgerald et al., 2009, C. Alos-Ferrer et al., 2012). We also predict that differences by frame will be observed in patterns of brain connectivity, following from results observed in the appetitive domain. Specifically, we expect to observe connectivity changes between the approach and avoidance conditions particularly in limbic regions such as the striatum, insula, and amygdala.

II. Methods

i. Participants

Fourteen healthy adult participants (9 female, mean age= 24.43, SD= 4.9) underwent functional MRI conducted at the Rutgers University Brain Imaging Center (RUBIC). Participants met standard MRI exclusion criteria (e.g.,. no metal implants, pregnancy, neurological disorders). Participants were recruited from the Rutgers University Newark community through

a department based subject recruitment system and word of mouth. Undergraduates were awarded course credit for participation. One participant was left handed. No participants reported taking medication for any psychiatric or neurological disorder. All participants gave written informed consent to participate. The study was approved by the Rutgers Institutional Review Board (protocol #12-530M).

Data from an independent cohort of subjects (n=14, 8 female, mean age =25.47, SD=4.37) was also used in analysis. This data was the subject of a previous manuscript (Mills-Finnerty et al., 2014). Subjects were screened based on the same criteria as the present study and were also scanned at the Rutgers University Brain Imaging Center. Participants in this cohort did not differ from the aversive framing cohort on age ($t(20.22) = 0.67651$, $p= 0.51$). Participant characteristics are described in more detail in Mills-Finnerty et al. (2014).

ii. Procedure

A version of the abstract reinforcer task (Mills-Finnerty et al., 2014) with aversive categories was developed through behavioral piloting with an independent group of subjects (n=49) to determine an appropriate range of categories, exemplars within those categories, and to optimize task format. Participants selected from a set of four categories: illnesses, car accidents, train incidents, and house incidents. A full list of category examples is available in Appendix A. Participants were asked to select the category they found the most negative. Participants unsure of how to select the most negative category were given the additional instruction to select the category with stimuli "they are most afraid of, or would least like to happen to them." Categories chosen as most negative by participants were car accidents (6), train incidents (5), and illnesses (3). No subjects chose house incidents. Each category contained 12 stimuli which all constituted

conditions that could lead to death (i.e. cancer, bomb threat on a train, house fire, head on car collision; refer to Appendix A).

In the scanner, participants made two-alternative forced choices between all possible combinations of category exemplars (i.e. "flu versus cancer"), once with the prompt "which would you rather avoid" (avoidance frame) and once as "which would you rather have" (approach frame). The scan run took 13 minutes and six seconds. Choices were presented in eight 28 second long blocks with 7 choices per block (except for the final block of each framing condition which contained 10 stimuli), for a total of 66 trials per framing condition and 132 trials total. Participants were given up to 4 seconds to respond, and after they selected their answer the screen changed to a crosshair to indicate the response had been logged. Twelve second rest periods divided the approach and avoidance blocks. Stimuli were presented and responses recorded using PsychoPy (http://www.psychopy.org/).

iii. Scanning Parameters

Functional imaging was conducted using a Siemens 3.0 Tesla Trio MRI scanner to acquire gradient echo T2*-weighted echo-planer (EPI) images with BOLD contrast. A 12 channel array coil was used due to increased signal detection in orbitofrontal regions. Each volume collected had 32 axial slices. 393 measurements were acquired in ascending contiguous order with a TR of 2s, for a total scan time of 13 minutes and 6 seconds. Imaging parameters included: field of view, 192 mm; slice thickness, 3mm; TR, 2s; TE, 30ms; flip angle, 90 degrees. Whole brain high resolution structural scans were acquired at 1 X 1 X 1 mm using an MP-RAGE pulse sequence.

iv. fMRI General Linear Model

Analysis was performed using FMRIB's Software Library (www.fmrib.ox.ac.uk/fsl). Skull stripping was performed using BET (Brain Extraction Tool) and then individual data was registered to the anatomical standard using FLIRT (FSL's Linear Registration Tool), in which the BOLD functional data are registered to the MPRAGE anatomical scan and then to the MNI atlas image. FEAT (FSL's Expert Analysis Tool) was used for all GLM analysis with the following parameters for first level (individual scan) analysis: motion correction with MCFLIRT; 5 mm FWHM spatial smoothing, highpass filtering using a value of 100s, and a second registration to the MNI atlas using 3 DOF. The two regressors used in first level analysis were the timepoints associated with the approach and avoidance frames; rest periods were used as baseline and therefore not modelled.

At the group level, activation was modelled several ways: as the average above baseline magnitude (activation) and below baseline magnitude (deactivation) of each framing condition (approach and avoidance); as a t test of the differences between activation in the approach and avoidance conditions; and the average group activation with approach and avoidance conditions collapsed together. This collapsing was done by modelling each subject's approach and avoidance related timepoints together in a first level analysis, producing individual files representing the average activation during both the approach and avoidance conditions, referred to here as the "all aversive" condition. All group models were run using the Flame 1 mixed effects model and corrected for multiple comparisons using a cluster threshold of $z=2.33$, $p>.05$. Head motion for the sample was minimal (<.5mm; mean=.26mm, SD=.14mm) and thus movement was not included as a regressor in group models. Motion did not differ between the aversive and appetitive framing subject cohorts, $t(25.9)= 0.24784$, $p= 0.81$. To ensure that null

effects using a z=2.33 cluster forming threshold were not false negatives, we also conducted several analyses using a cluster forming threshold of z=1.65

In order to further clarify how magnitude increases and decreases differ based on valence of stimuli, data from the appetitive framing task reported in Mills-Finnerty et al. (2014) was compared directly to the aversive framing data from the present study. In Mills-Finnerty et al. (2014) participants completed a task with the same format as in the present study, except the individualized categories of stimuli were appetitive (vacation destinations, leisure activities, etc.) and the choice framing was either positive ("which do you like more") or negative ("which do you like less;" refer to Mills-Finnerty et al., 2014 for more detailed methods and participant information). Positive appetitive framing ("which do you like more") was compared to avoidance aversive framing ("which would you rather avoid"), and negative appetitive framing ("which do you like less") was compared to approach framing for aversive stimuli ("which would you rather have") using t-tests to measure differences in activation magnitude between these conditions.

v. Connectivity

Connectivity analysis was performed to quantify how brain network response during decisions for abstract aversive reinforcers is influenced by framing. While general linear model analysis addresses how conditions can affect the level of response by various brain regions, it can not reveal how those brain regions interact. Here, we use an Independent Multi-sample Greedy Equivalence Search (IMaGES). The algorithm starts with an empty graph and searches forward, one new connection at a time, until it finds the set of connections that optimally represents the entire group of subjects, interpolating any missing data. The algorithm searches with the restriction of finding only Markov equivalence classes of directed acyclic graphs. The process is

penalized to prevent overfitting using the Bayes Information Criterion (Schwarz, 1978): -2ln(ML) + k ln(n), where ML is the maximum likelihood estimate, k is the dimension of the model (the number of directed edges plus the number of variables), and n is the sample size (number of participants). The LOFS post search filter was used to orient the direction of connections. LOFS "exploits the fact that the residuals of the correct linear model with independent non-Gaussian errors will be less Gaussian that the residuals of any incorrect model. That can be seen from two facts: (1) a sum of i.i.d. non-Gaussian variables is (usually) closer to Normal than any of the terms in the sum; and (2) the regression residual of a variable X on a false orientation of its adjacent variables is a weighted sum of the error term for X and the error terms for the variables of mis-oriented edges—whereas on the correct orientation the residual for X is just the error term for X" (Ramsey et al., 2011). Edge orientation should be interpreted as a summary of the dominant direction of an edge, with the assumption that in biological reality communication likely volleys back and forth between brain regions in many cases. Orienting edges to be unidirectional rather than bidirectional is done here for the sake of improving model precision and recall based on simulation results (Ramsey et al., 2011), as well as recent empirical validations that this method correctly identifies 'ground truth" directionality, in experimental conditions where this information is known (Mill et al., 2016).

ROIs were chosen based on activation during GLM analysis. Main effects used to generate coordinates for region of interests used in connectivity analysis were also validated using threshold-free cluster enhancement (see Supplemental Methods). Binary masks were created for VMPFC and bilateral putamen using FSL view and the Harvard-Oxford anatomical atlas, in which the probabilistic atlas defined ROIs were converted into masks. Since activation both above and below baseline were observed using GLM analysis, regions where both

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

activations and deactivations occurred were masked using more conservative methods. Specifically, the hippocampus mask was thresholded to 70% anatomical probability to exclude activation likely to be situated in other regions. For the insula, anterior cingulate, and amygdala, coordinates were restricted to those that fell within <70% probability of being a part of that region, and were then selected using the center of the clusters active or deactive identified using group GLM analysis. A 9mm sphere was then created to mask that activation. Mask coordinates were chosen to ensure minimal overlap of active and deactive voxels and are listed in Table 1. For the insula, two masks were created to account for both activations and deactivation, one in anterior insula (activation) and one in posterior (deactivation). No voxel overlap occurred between the anterior cingulate, hippocampus, or amygdala masks, and minimal overlap (approx. 3 voxels) was observed for the insula and hippocampus masks.

Average time series for each subject were extracted from these ROIs using FSL's meanTS module. The first and last TR of all condition blocks after the first block were excluded from analysis to exclude any carry over effects resulting from the hemodynamic response function time lag. Time courses of interest were arranged into a matrix for each subject, with the ROIs as columns and each row representing a single time point. These files were then input into the IMaGES workflow in Tetrad. IMaGES outputs a set of graphs that are all equivalently likely called a Markov Equivalence Class (MEC). Final graphs were selected by choosing the most complex graph (the one with the most edges) within the MEC generated for each condition. Edge (connection) weights were exported from Tetrad into LibreOffice Calc (https://www.libreoffice.org/). T statistics were averaged across the group, and were used instead of raw coefficient values because they take into account standard error. The TDIST function was

used to calculate significance values of graph edges. Graph structure was input into Cytoscape (www.cytoscape.org) for visualization and calculation of graph metrics.

ROI Mask	MNI coordinates		
	x	y	z
Anterior cingulate	45	72	55
Left anterior insula	64	71	34
Right anterior insula	26	71	34
Left posterior insula	65	58	34
Right posterior insula	25	58	34
Left amygdala	57	62	23
Right amgydala	33	62	23
mPFC	45	79.2	28
Right putamen	31	63.5	36.2
Left putamen	58	63.5	36.2
Left hippocampus	58	53.3	27.8
Right hippocampus	31	53.3	27.8

Table 1. MNI coordinates used to define region of interest masks.

III. Results

i. Behavioral

Reaction time was significantly longer for the approach (M=2.16, SD=.29) compared to the avoidance condition (M=1.99, SD=.34; t = -6.3812, df = 13, p=.00002).

ii. fMRI

Greater activation was observed for the contrast of the avoidance frame>approach frame in the right insula, right postcentral gyrus, and bilateral caudate using an Ordinary Least Squares regression with a cluster threshold of z=2.33, p<.05 (Figure 1, top). No activation was greater during the approach frame when compared to the avoidance frame using a cluster threshold of z=1.65, p=.05. Significant activation was observed for the "all aversive" condition (collapsed across framing conditions), in the right dorsal caudate, bilateral thalamus, pre- and postcentral

gyrus, supplementary motor area, anterior cingulate, lateral occipital cortex, superior parietal lobule, angular gyrus, middle temporal gyrus, and left hippocampus at a cluster threshold of z=2.33, p<.05 (Figure 1, bottom).

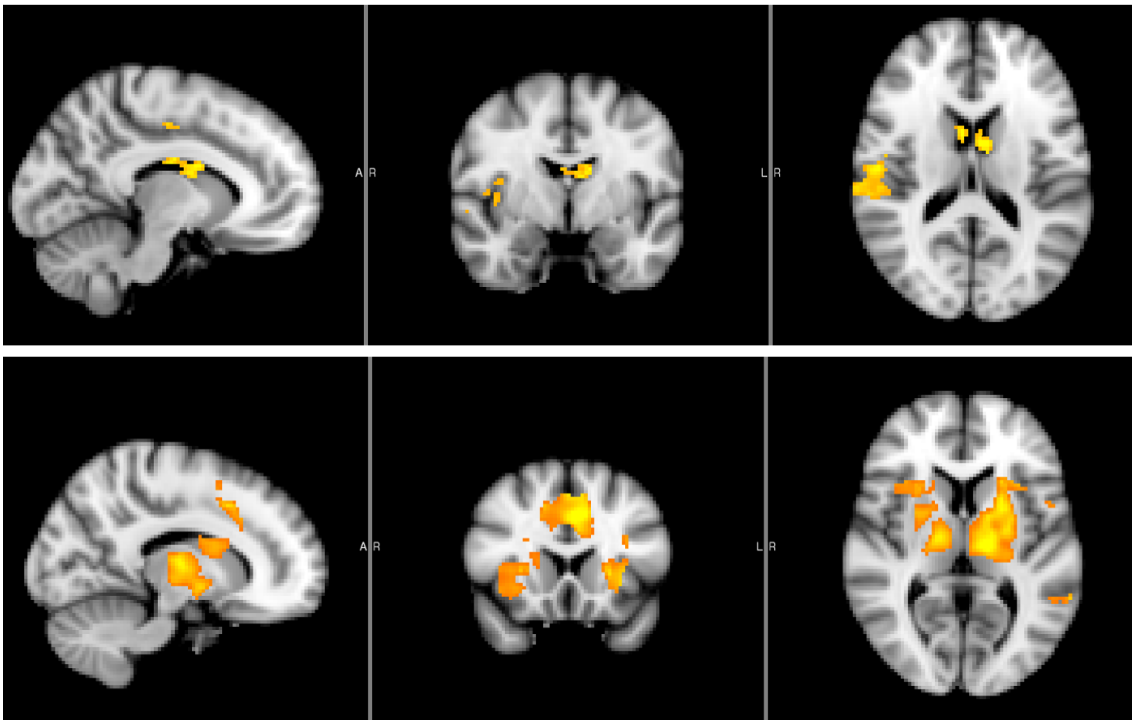


Figure 1. Top, activation greater for avoidance > approach framed choices. Bottom, significant activation during all choices (collapsed across frame).

Significant deactivations were also observed for the all aversive condition, in the right insula, VMPFC, posterior cingulate, superior parietal lobule, right supramarginal gyrus, and right postcentral gyrus at a cluster threshold of z=3, p<.05 (Figure 2, pictured using a cluster threshold of z=2.33, p=.05 for visualization purposes). Deactive regions largely overlapped between the aversive and approach conditions, with the exception of clusters in right thalamus and posterior cingulate during the avoidance frame, and in superior temporal gyrus in the approach frame. More information about significant activation cluster location can be found in Table 2.

Condition	Cluster	Voxels	P	Z-MAX	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)
Avoidance > Approach Framing	3	871	0.00000435	4.33	54	-28	14
	2	345	0.00845	4.14	-4	-30	56
	1	259	0.039	4.06	-6	-2	16
Activation, all aversive choices	5	8619	6.37E-23	6.24	-36	-14	60
	4	2105	0.0000000193	4.93	14	-14	8
	3	2038	0.0000000596	5.31	-26	-64	48
	2	667	0.0018	3.63	26	-56	52
	1	409	0.0282	4.23	-48	-40	-6
Deactivation, all aversive choices	5	2428	0.00000000069	4.34	8	-30	40
	4	1565	0.000000358	4.24	40	-14	-8
	3	1146	0.000012	4.58	6	6	-12
	2	976	0.0000551	4.15	-28	-24	-22
	1	935	0.0000807	3.99	46	-34	28

Table 2. Cluster extent, significance, and location for condition contrasts and main effects of task.

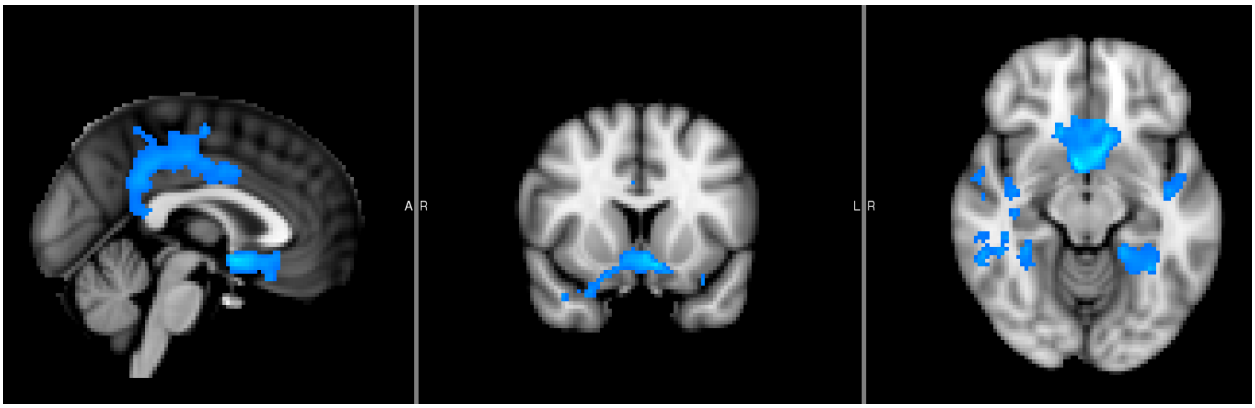


Figure 2. Regions showing significant decreases in activation ("deactivation") relative to baseline for all choices (collapsed across frame).

Data from the appetitive framing task reported in Mills-Finnerty et al. (2014) was compared directly to the data in the present study. For all contrasts of aversive framing>appetitive framing, no activation was observed above a cluster threshold of z=1.65, p=.05. For the direct contrast of appetitive framing > aversive framing, activation was observed that differed by frame. Specifically, for positive appetitive framing ("which do you like more") compared to avoidance aversive framing ("which would you rather avoid"), greater activation was observed in bilateral insula, anterior and posterior cingulate, precuneus, and bilateral lateral

occipital cortex at a cluster threshold of z=1.65, p=.05. Since the goal of this analysis was simply to confirm that appetitive choice response involves mainly increases in activation while aversive choice involves decreases, we feel this non-null result is of interest to report although it does not meet the more stringent criteria using z=2.33, p=.05. For negative appetitive framing ("which do you like less") compared to approach framing for aversive stimuli ("which would you rather have"), greater activation was observed in VMPFC at a cluster threshold of z=2.33, p=.05.

iii. Connectivity

Approach and avoidance related connectivity was measured separately in the following network of regions: putamen, anterior insula, and anterior cingulate (areas active above baseline); and posterior insula, VMPFC, hippocampus, and amygdala (areas active below baseline). A connection to B, originating from A, is indicated here as A->B, whereas a connection from B to A is indicated as B->A. During both avoidance and approach framing, the following connections were observed: putamen->anterior insula, putamen->anterior cingulate, putamen->posterior insula, putamen->hippocampus, VMPFC->amygdala, amygdala->hippocampus (Figure 3). During avoidance framing, additional connections were observed from posterior insula->amygdala and putamen->VMPFC. For all connections, the probability of them occurring by chance measured against a t distribution was p<.0005.

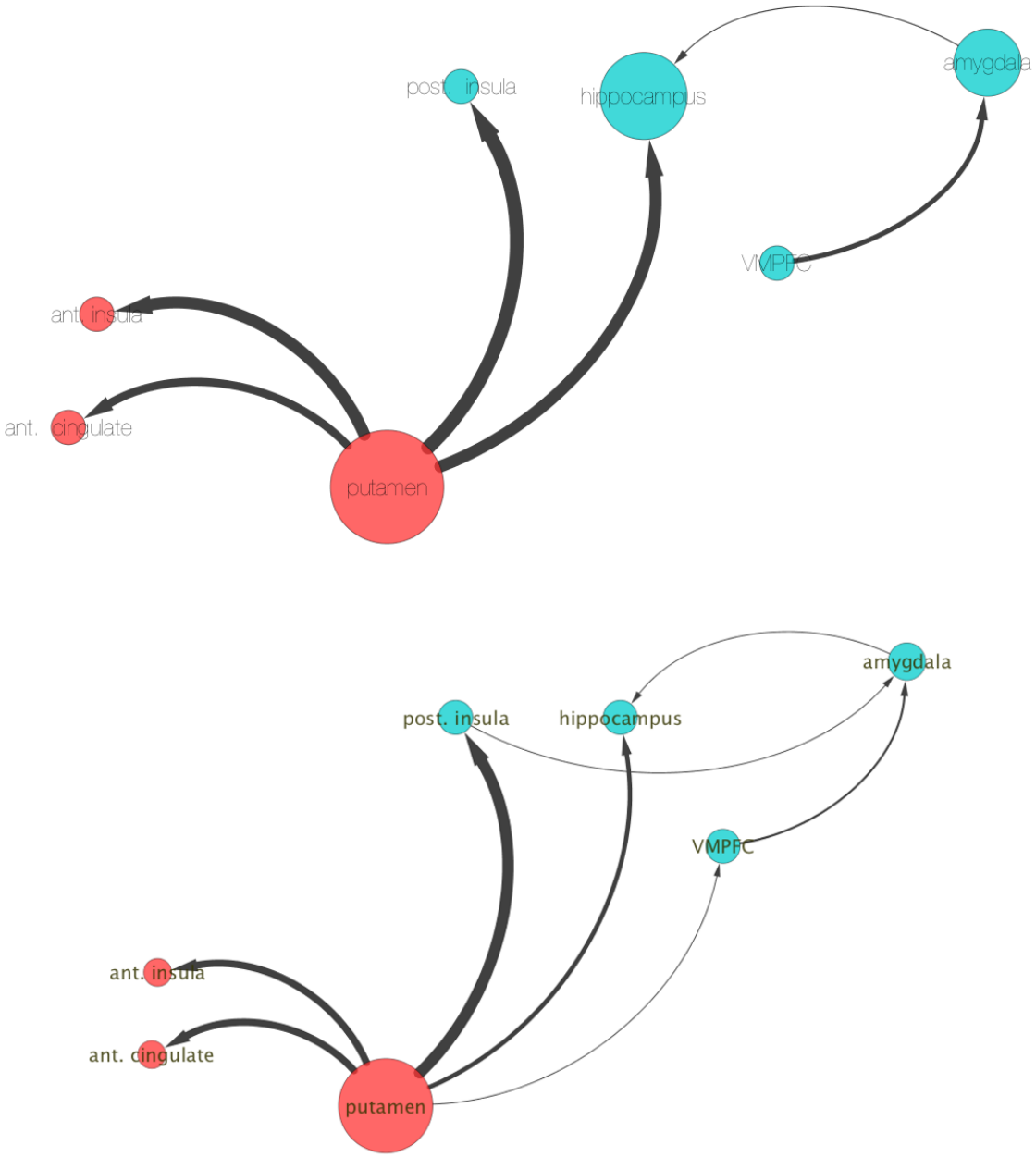


Figure 3. Graph models of connectivity during approach framing (top) and avoidance framing (bottom).

IV. Discussion

The present study characterizes brain response to aversive hypothetical stimuli framed as approach or avoidance choices. Widespread, robust deactivation was observed within regions associated with decision making during choices for aversive stimuli, and sensitivity to choice context (approaching vs. avoiding an aversive stimulus) was observed via increases in limbic connectivity amongst deactive regions during avoidance choices. Taken together these findings suggest that the BOLD response to aversive abstract reinforcers involves primarily deactivation, suggesting valence sensitivity in deactive regions. The hypothesis that choosing which aversive stimulus to avoid would be processed similarly to choosing which appetitive stimulus to approach was supported, via greater activation and network connectivity for avoidance>approach framing within reward sensitive regions.

i. Framing and aversive abstract reinforcer response magnitude

Framing effects have been robustly observed in the concrete context (e.g. Kahneman & Tversky 1986; refer to Kuhberger 1998 for meta-analysis), and recently established in the abstract context as well (Foo et al., 2014; Mills-Finnerty et. al., 2014). However, no studies to our knowledge have tested the effects of approach and avoidance frames on choices for hypothetical aversive abstract reinforcers. Consistent with predictions and the existing literature (e.g. C. Alos-Ferrer et. al., 2012, Foo et. al., 2014), significant reaction time differences were observed for approach versus avoidance frames, with significantly faster RT for avoidance compared to approach. Since RT is typically interpreted as an index of task difficulty, these results suggest that choosing which aversive reinforcer to approach is more difficult than choosing which one to avoid. Since avoidance is the more positive or desirable outcome, it follows that these choices can be made more quickly and easily.

Despite highly significant differences in reaction time for approach and avoidance framing, no differences in brain activation were observed for the direct contrast of approach>avoidance choices. For the contrast of avoidance>approach activity was observed in the caudate, insula, and post-central gyrus. Average activation for approach and avoidance largely occurred in overlapping regions. Thus, it appears that the framing manipulation has smaller effects on magnitude *increases* in the context of aversive choices. Connectivity analysis results suggest that there are instead significant effects that occur via *decreases* in activation, in contrast to results in the appetitive domain (Mills-Finnerty et al., 2014). These results suggest a valence sensitive account of processing of hypothetical aversive choices.

ii. Connectivity dynamics underlying framing effects in the aversive domain

In contrast to the GLM results, effects of approach vs. avoidance frame were observed via connectivity analysis and shed light on differences between processing of appetitive and aversive abstract reinforcers. Many of the areas that showed greater activation during appetitive framing in previous studies exhibited significant deactivation during aversive framing, including the insula and mPFC. Results from connectivity analysis suggest frame-based differences in deactivation.

The putamen appears to play a central role in both activation and deactivation networks during both the approach and avoidance frames. There were more connections between the putamen and several deactive regions (posterior insula and VMPFC) during avoidance, but not approach framing. The putamen had the most connections of any region in the network and highest betweenness centrality (BC) score during both conditions. BC is an index of how many of the shortest paths in a network pass through that node and indicates that the putamen is highly central to the graph. Results from the literature suggest that aversive prediction error responses

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

are coded by regions of caudate and putamen (e.g. Gottfreid et al., 2002: O'Doherty et al., 2006; Delgado et al., 2008; see Bissonnette et al., 2014 for review). Several studies have used both appetitive and aversive stimuli to measure PE. For example, one study found that the putamen, in addition to the anterior insula and rostral anterior cingulate, was responsive during prediction errors involving both unexpected relief and exacerbation of pain (Seymour et al., 2005). Interestingly, the specific sub-regions of the striatum, insula, and anterior cingulate that decreased activation in response to prediction error in Seymour et al. (2005) were active in our study, whereas the posterior insula and posterior cingulate both contained deactive voxels. In another study that used high resolution imaging (Mattfield et. al., 2011), the region of caudate that is active for positive PE (right caudate head) is deactive during the all aversive aversive condition in our results. The more anterior portion of the caudate that showed greater deactivation during negative PE in their study had greater activation in ours. These results suggest that the same regions that are involved more generally in PE are active or deactive during our task. However, without high resolution imaging and given the differences in protocols, it is difficult to interpret how meaningful differences in voxel cluster location are, or how much of the difference in effects is due to the use of real versus hypothetical rewards. Further, since there are no expectations or actual outcomes in our task, it is unlikely that putamen activation or connectivity represents prediction error. It is possible that the putamen codes the hypothetical outcomes associated with choices, resulting in relative increases in activation when avoiding an aversive stimulus. To better clarify value and salience dynamics, in future studies participants could explicitly rate each of these factors, ideally after every choice. However, the primarily deactivation-based dynamics observed provide support for valence sensitive processes during choices for aversive abstract reinforcers.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

The involvement of the anterior cingulate via activation increases and connectivity with the putamen may reflect its role in conflict-based decision making. Avoiding and approaching aversive stimuli both involve forced choices between stimuli that are both highly aversive, a context inducing decision conflict. The anterior cingulate has been implicated in decision conflict, playing a role in information integration and control signaling during choices resulting in losses, by optimizing strategies to minimize loss (Brown & Alexander, 2014), such as by coding "teaching signals" used to inform avoidance learning (Botvinick, 2007). It is unclear what optimization strategies participants may have used to weigh aversive choice options, for example by adaptively learning choice heuristics throughout the course of the task (e.g. "always avoid cancer"). Future studies designed to investigate such potential individual differences are needed to clarify the role of anterior cingulate more specifically. Since in this task there are no outcomes to influence, it is possible the ACC plays more of an integration role in consolidating information to resolve decision conflicts, which is consistent with the similar strength of connectivity between ACC-putamen and same direction of influence in both framing conditions. The striatum has also been implicated in choice conflict, responding based on degree of cognitive control (rather than effort) during attentional interference (Robertson et al., 2015). Optogenetic manipulation of circuits targeting striatal striosomes in animal models revealed that cost-benefit choices, but not benefit-benefit or cost-cost choices, can be manipulated in particular cell populations (Friedman et al., 2015), suggesting strong interactions between decision context and striatal function. Anterior cingulate and putamen activation, connectivity strength, and direction of connection did not differ significantly by frame, suggesting a similar response to choice conflict in both framing contexts.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

The deactive regions in the network had more intra-connection than the active regions in both framing conditions. This deactivation network connectivity increased substantially during avoidance framing, with two unique connections (putamen->VMPFC, posterior insula->amygdala). This increase in deactive network connectivity in limbic regions for avoidance compared to approach is in line with predictions regarding the brain response to avoiding a negative stimulus. Specifically, it was predicted that areas such as the putamen and mPFC which increase activation during positively framed choices for appetitive abstract reinforcers should behave similarly given a choice to avoid an aversive abstract reinforcer. This prediction was partially confirmed, in that putamen increased its activation for avoidance>approach frames, but mPFC decreased its activation. Connectivity between mPFC and putamen increased during avoidance framing, suggesting that that the decreases in mPFC during avoidance framing may actually be driven directly by the increases in putamen activation. The putamen may code factors such as the hypothetical aversiveness of the choice options, information that may be incorporated into a value signal in mPFC.

The presence of activation and deactivation within different sub-regions of the same brain areas also suggests that potentially opponent processes are co-occuring in response to aversive stimuli. This delineation may be based on functional specializations of these subregions. For example, activation was observed in the anterior insula and deactivation in the posterior insula. These sub-regions have been implicated in different aspects of interoception - anterior insula with cognitive and affective components (such as feelings of disgust) and posterior insula with sensory encoding (such as the experience of pain; see review by Uddin, 2014). Interestingly, connectivity analysis revealed connections between the putamen and both anterior and posterior insula during both approach and avoiding framing. During avoidance framing only, an additional

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

connection from posterior insula to the amgydala was also present. These results suggest that posterior insula is the sub-region more affected by the difference between approach and avoidance prompts for aversive stimuli. Given the role of the amygdala in responding to aversive stimuli (e.g. O'Doherty, 2001; Whalen et al., 2004; Orsini et al., 2015), particularly during fear learning (e.g. Nader et al., 2000; Wolff et al., 2014; Moscarello et al., 2014) and in relation to loss aversion (e.g.DeMartino et al., 2010; Tremblay et al., 2014, Canessa et al., 2013), these results suggest that inputs from the posterior insula may directly influence this response, such as by relaying information about relevant sensory features of hypothetical choices (such as the feeling of symptoms associated with different illnesses).

iii. Appetitive vs. Aversive framing effects

To further clarify valence effects on choices for hypothetical stimuli, choices for hypothetical appetitive stimuli were compared to similarly framed choices for aversive stimuli. Specifically, positive appetitive framing ("which do you like more") was compared to avoidance aversive framing ("which would you rather avoid"), while negative appetitive framing ("which do you like less") was compared to approach framing for aversive stimuli ("which would you rather have"). Greater activation was observed for contrasts of appetitive>aversive framing, but not for any contrasts of aversive>appetitive framing, suggesting that valence influences choices for abstract reinforcers by leading predominantly increases when the ARs are appetitive. Connectivity modelling results suggest that interactions amongst active regions change based on frame in the appetitive domain, whereas changes in deactive region connectivity drives a significant amount of frame-based responding in the aversive domain. It is of course possible that dimensions other than valence may drive the difference in magnitude based response between tasks, such as differing sensory elements of choices, or different mechanisms for

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

computing appetitive vs. aversive value, and further studies will be needed to fully clarify these differences.

Additionally, conflicting results in the literature in support of the salience and valence accounts may be due to protocol differences, such as contextual changes (gambling vs. certain choices, learning vs. passive tasks, etc.) that may drive responding to be more activation or deactivation based. Here, the appetitive and aversive choice protocols were visually highly similar and subjects were scanned using the same scanner, however the aforementioned differences in stimuli do limit the inferences that can be made from this comparison. To help resolve this, future analyses could use measures such as percent signal change to characterize the activation increases and decreases in each condition in a within-subjects design, in particular to determine if areas such as mPFC increase or decrease activation in a manner that is parametrically related to increase and decreases in stimulus value. Measuring physiological reactions to stimuli would also help bolster inferences about how individual differences in emotional responding or arousal might mediate connectivity patterns. The primary limitation of the present study is the small sample size, and future studies replicating these results with a larger sample are needed for several reasons. Although the strong behavioral effect of decision frame reported in Mills-Finnerty (2014) replicated using a different stimulus set in the present study, our sample size precludes an investigation of individual differences related to gender, handedness, or other potential variables that might be related to decision making biases (e.g. numeracy). Although the edges in our connectivity model were all significant with sample size included in the DOF of the IMaGES model, it will be important to replicate these effects with a larger sample size.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3340v1 | CC BY 4.0 Open Access | rec: 12 Oct 2017, publ: 12 Oct 2017

In sum, we demonstrate that choices for hypothetical aversive choices rely on similar brain substrates as those involved in appetitive hypothetical choice. Activation patterns involved both increases and decreases in magnitude, suggesting that brain response is sensitive to valence in this context. Further, approach and avoidance frames seem to differentially modulate activation, with differences primarily observed via connectivity changes among deactive regions. These results provide a novel characterization of how network communication patterns among both active and deactive regions shift based on stimulus valence and choice framing.

ACKNOWLEDGEMENTS: This work was supported by a grant from the James S. McDonnell Foundation, and was conducted at Rutgers University.

DISCLOSURES: No competing financial interests exist on the part of the authors.

References

Alos-Ferrer, C., & Shi, F. (2012). Choice-Induced Preference Change: In Defense of the Free-Choice Paradigm. *Available at SSRN 2062507*, (February). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2062507

Benoit, R. G., Szpunar, K. K., & Schacter, D. L. (2014). Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(46), 16550–5. http://doi.org/10.1073/pnas.1419274111

Bhanji, J. P., & Delgado, M. R. (2014). Perceived control influences neural responses to setbacks and promotes persistence. *Neuron*, *83*(6), 1369–75. http://doi.org/10.1016/j.neuron.2014.08.012

Bissonette, G. B., Gentry, R. N., Padmala, S., Pessoa, L., & Roesch, M. R. (2014). Impact of appetitive and aversive outcomes on brain responses: linking the animal and human literatures. *Frontiers in Systems Neuroscience*, *8*(March), 24. http://doi.org/10.3389/fnsys.2014.00024

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, Affective & Behavioral Neuroscience*, *7*(4), 356–366. http://doi.org/10.3758/CABN.7.4.356

Brooks, A. M., & Berns, G. S. (2013). Aversive stimuli and loss in the mesocorticolimbic dopamine system. *Trends in Cognitive Sciences*, *17*(6), 281–286. http://doi.org/10.1016/j.tics.2013.04.001

Brown, Joshua W., Alexander, W. H. (2013). Foraging Value, Risk Avoidance, and Multiple Control Signals: How the Anterior Cingulate Cortex Controls Value-based

Decision-making. *Journal of Cognitive Neuroscience*, *26*(3), 194–198. http://doi.org/10.1162/jocn

Canessa, N., Crespi, C., Motterlini, M., Baud-bovy, G., Chierchia, G., Pantaleo, G., … Cappa, S. F. (2013). The Functional and Structural Neural Basis of Individual Differences in Loss Aversion, *33*(36), 14307–14317. http://doi.org/10.1523/JNEUROSCI.0497-13.2013

Caria, A., Sitaram, R., Veit, R., Begliomini, C., & Birbaumer, N. (2010). Volitional control of anterior insula activity modulates the response to aversive stimuli. A real-time functional magnetic resonance imaging study. *Biological Psychiatry*, *68*(5), 425–32. http://doi.org/10.1016/j.biopsych.2010.04.020

Cavallo, A., Heyes, C., Becchio, C., Bird, G., & Catmur, C. (2014). Timecourse of mirror and counter-mirror effects measured with transcranial magnetic stimulation. *Social Cognitive and Affective Neuroscience*, *9*, 1082–1088. http://doi.org/10.1093/scan/nsu085

Chikazoe, J., Lee, D. H., Kriegeskorte, N., & Anderson, A. K. (2014). Population coding of affect across stimuli, modalities and individuals. *Nature Neuroscience*, *17*(8), 1114–1122. http://doi.org/10.1038/nn.3749

Chiu, Y., Cools, R., & Aron, A. (2014). Opposing Effects of Appetitive and Aversive Cues on Go/No-go Behavior and Motor Excitability, 1–10. http://doi.org/10.1162/jocn

Cohen, J. (1962). the Statistical Power of Abnormal-Social Psychological Research : a Review. *Journal of Abnormal and Social Psychology*, *65*(3), 145–153. http://doi.org/10.1037/h0045186

Collins, K. a., Mendelsohn, A., Cain, C. K., & Schiller, D. (2014). Taking Action in the Face of Threat: Neural Synchronization Predicts Adaptive Coping. *Journal of Neuroscience*, *34*(44), 14733–14738. http://doi.org/10.1523/JNEUROSCI.2152-14.2014

Cooper, J. C., & Knutson, B. (2008). Valence and salience contribute to nucleus accumbens activation. *NeuroImage*, *39*(1), 538–47. http://doi.org/10.1016/j.neuroimage.2007.08.009

Cunningham, W. a., & Brosch, T. (2012). Motivational Salience: Amygdala Tuning From Traits, Needs, Values, and Goals. *Current Directions in Psychological Science*, *21*(1), 54–59. http://doi.org/10.1177/0963721411430832

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science (New York, N.Y.)*, *313*(5787), 684–687. http://doi.org/10.1126/science.1128356

Delgado, M. R., Locke, H. M., Stenger, V. a, & Fiez, J. a. (2003). Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations. *Cognitive, Affective & Behavioral Neuroscience*, *3*(1), 27–38. http://doi.org/10.3758/CABN.3.1.27

Delgado, M. R., Li, J., Schiller, D., & Phelps, E. a. (2008). The role of the striatum in aversive learning and aversive prediction errors. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1511), 3787–800. http://doi.org/10.1098/rstb.2008.0161

Dixon, M. L., & Christoff, K. (2014). The lateral prefrontal cortex and complex value-based learning and decision making. *Neuroscience and Biobehavioral Reviews*, *45*, 9–18. http://doi.org/10.1016/j.neubiorev.2014.04.011

Durnez, J., Degryse, J., Moerkerke, B., Seurinck, R., Sochat, V., Poldrack, R., & Nichols, T. (2016). Power and sample size calculations for fMRI studies based on the prevalence of active peaks. *bioRxiv*, 49429. http://doi.org/10.1101/049429

Feldman Hall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience*, *7*, 743–751. http://doi.org/10.1093/scan/nss069

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S., & Mobbs, D. (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience*, *7*(7), 743–51. http://doi.org/10.1093/scan/nss069

Foo, J. C., Haji, T., & Sakai, K. (2014). Prefrontal Mechanisms in Preference and Non-Preference-based Judgments. *NeuroImage*. http://doi.org/10.1016/j.neuroimage.2014.03.046

Friedman, A., Homma, D., Gibb, L. G., Amemori, K. I., Rubin, S. J., Hood, A. S., … Graybiel, A. M. (2015). A corticostriatal path targeting striosomes controls decision-making under conflict. *Cell*, *161*(6), 1320–1333. http://doi.org/10.1016/j.cell.2015.04.049

Gerlach, K. D., Spreng, R. N., Gilmore, A. W., & Schacter, D. L. (2011). Solving future problems: default network and executive activity associated with goal-directed mental

simulations. *NeuroImage*, *55*(4), 1816–24. http://doi.org/10.1016/j.neuroimage.2011.01.030

Gilaie-Dotan, S., Tymula, a., Cooper, N., Kable, J. W., Glimcher, P. W., & Levy, I. (2014). Neuroanatomy Predicts Individual Risk Attitudes. *Journal of Neuroscience*, *34*(37), 12394–12401. http://doi.org/10.1523/JNEUROSCI.1600-14.2014

Gläscher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., … Tranel, D. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(36), 14681–6. http://doi.org/10.1073/pnas.1206608109

Gottfried, J. a, O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *22*(24), 10829–37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12486176

Gu, X., Liu, X., Van Dam, N. T., Hof, P. R., & Fan, J. (2013). Cognition-emotion integration in the anterior insular cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(1), 20–7. http://doi.org/10.1093/cercor/bhr367

Harris, A., Adolphs, R., Camerer, C., & Rangel, A. (2011). Dynamic construction of stimulus values in the ventromedial prefrontal cortex. *PloS One*, *6*(6), e21074. http://doi.org/10.1371/journal.pone.0021074

Hayes, D. J., Duncan, N. W., Xu, J., & Northoff, G. (2014). A comparison of neural responses to appetitive and aversive stimuli in humans and other mammals. *Neuroscience and Biobehavioral Reviews*, *45*, 350–68. http://doi.org/10.1016/j.neubiorev.2014.06.018

Hayes, D. J., & Northoff, G. (2011). Identifying a network of brain regions involved in aversion-related processing: a cross-species translational investigation. *Frontiers in Integrative Neuroscience*, *5*(October), 49. http://doi.org/10.3389/fnint.2011.00049

Hayes, D. J., & Northoff, G. (2012). Common brain activations for painful and non-painful aversive stimuli. *BMC Neuroscience*, *13*, 60. http://doi.org/10.1186/1471-2202-13-60

J. Silvers, T. Wager, J Weber, K. O. (2014). The neural basis of uninstructed negative emotion regulation. *Social Cognitive and Affective Neuroscience*.

Jacob Westfall 1† Thomas E. Nichols 2 Tal Yarkoni 1†. (2016). Fixing the stimulus as fixed effect fallacy in task fMRI.

Jensen, J., Smith, A. J., Willeit, M., Crawley, A. P., Mikulis, D. J., Vitcu, I., & Kapur, S. (2007). Separate brain regions code for salience vs. valence during reward prediction in humans. *Human Brain Mapping*, *28*(4), 294–302. http://doi.org/10.1002/hbm.20274

John, J. P., Halahalli, H. N., Vasudev, M. K., Jayakumar, P. N., & Jain, S. (2011). Regional brain activation/deactivation during word generation in schizophrenia: fMRI study. *British Journal of Psychiatry*, *198*, 213–222. http://doi.org/10.1192/bjp.bp.110.083501

Joshua, M., Adler, A., Mitelman, R., Vaadia, E., & Bergman, H. (2008). Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(45), 11673–84. http://doi.org/10.1523/JNEUROSCI.3839-08.2008

Kahnt, T., Park, S. Q., Haynes, J.-D., & Tobler, P. N. (2014). Disentangling neural representations of value and salience in the human brain. *Proceedings of the National*

Academy of Sciences of the United States of America, *111*(13), 5000–5. http://doi.org/10.1073/pnas.1320189111

Kang, M. J., & Camerer, C. F. (2013). fMRI evidence of a hot-cold empathy gap in hypothetical and real aversive choices. *Frontiers in Neuroscience*, *7*(June), 104. http://doi.org/10.3389/fnins.2013.00104

Kessler, D., Angstadt, M., & Sripada, C. S. (2017). Reevaluating "cluster failure" in fMRI using nonparametric control of the false discovery rate. *Proceedings of the National Academy of Sciences*, *114*(17), E3372–E3373. http://doi.org/10.1073/pnas.1614502114

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biology*, *4*(8), e233. http://doi.org/10.1371/journal.pbio.0040233

Kirlic, N., Young, J., & Aupperle, R. L. (2016). Animal to human translational paradigms relevant for approach avoidance conflict decision making. *Behaviour Research and Therapy*. http://doi.org/10.1016/j.brat.2017.04.010

Kolling, N., Wittmann, M., & Rushworth, M. F. S. (2014). Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron*, *81*(5), 1190–1202. http://doi.org/10.1016/j.neuron.2014.01.033

Kuhberger. (1998). The Influence of Framing on Risky Decisions :, *75*(1), 23–55.

Lamm, C., Silani, G., & Singer, T. (2015). Distinct neural networks underlying empathy for pleasant and unpleasant touch. *Cortex*. http://doi.org/10.1016/j.cortex.2015.01.021

Lammel, S., Ion, D. I., Roeper, J., & Malenka, R. C. (2011). Projection-Specific Modulation of Dopamine Neuron Synapses by Aversive and Rewarding Stimuli. *Neuron*, *70*(5), 855–862. http://doi.org/10.1016/j.neuron.2011.03.025

Lawson, R. P., Seymour, B., Loh, E., Lutti, A., Dolan, R. J., Dayan, P., … Roiser, J. P. (2014). The habenula encodes negative motivational value associated with primary punishment in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(32), 11858–63. http://doi.org/10.1073/pnas.1323586111

Li, S., Jiang, X., Yu, H., & Zhou, X. (2014). Cognitive empathy modulates the processing of pragmatic constraints during sentence comprehension. *Social Cognitive and Affective Neuroscience*, *9*, 1166–1174. http://doi.org/10.1093/scan/nsu091

Lindquist, K. a, Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2015). The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature. *Cerebral Cortex (New York, N.Y. : 1991)*, 1–13. http://doi.org/10.1093/cercor/bhv001

Liu, X., Hairston, J., Schrier, M., & Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: A meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1219–1236. http://doi.org/10.1016/j.neubiorev.2010.12.012

Ma, N., Baetens, K., Vandekerckhove, M., Van der Cruyssen, L., & Van Overwalle, F. (2014). Dissociation of a trait and a valence representation in the mPFC. *Social Cognitive and Affective Neuroscience*, *9*(10), 1506–14. http://doi.org/10.1093/scan/nst143

Martín-Loeches, M., Hernández-Tamames, J. a, Martín, a, & Urrutia, M. (2014). Beauty and ugliness in the bodies and faces of others: an fMRI study of person esthetic judgement. *Neuroscience*, *277*, 486–97. http://doi.org/10.1016/j.neuroscience.2014.07.040

Mattfeld, a. T., Gluck, M. a., & Stark, C. E. L. (2011). Functional specialization within the striatum along both the dorsal/ventral and anterior/posterior axes during associative learning via reward and punishment. *Learning & Memory*, *18*, 703–711. http://doi.org/10.1101/lm.022889.111

Metereau, E., & Dreher, J.-C. (2014). The medial orbitofrontal cortex encodes a general unsigned value signal during anticipation of both appetitive and aversive events. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *63C*, 42–54. http://doi.org/10.1016/j.cortex.2014.08.012

Mill, R. D., Bagic, A., Bostan, A., & Cole, M. W. (2016). Empirical validation of directed functional connectivity. *NeuroImage*. http://doi.org/10.1016/j.neuroimage.2016.11.037

Mills-Finnerty, C., Hanson, C., & Hanson, S. J. (2014). Brain network response underlying decisions about abstract reinforcers. *NeuroImage*. http://doi.org/10.1016/j.neuroimage.2014.09.019

Moscarello, J. M., & LeDoux, J. E. (2013). Active Avoidance Learning Requires Prefrontal Suppression of Amygdala-Mediated Defensive Reactions. *Journal of Neuroscience*, *33*(9), 3815–3823. http://doi.org/10.1523/JNEUROSCI.2596-12.2013

Mosher, C. P., & Rudebeck, P. H. (2015). news and views The amygdala accountant : new tricks for an old structure. *Nature Publishing Group*, *18*(3), 324–325. http://doi.org/10.1038/nn.3949

Mumford, J. A. (2012). A power calculation guide for FMRI studies. *Social Cognitive and Affective Neuroscience*, *7*(6), 738–742. http://doi.org/10.1093/scan/nss059

Murch, K. B., & Krawczyk, D. C. (2014). A neuroimaging investigation of attribute framing and individual differences. *Social Cognitive and Affective Neuroscience*, *9*(10), 1464–71. http://doi.org/10.1093/scan/nst140

Nader, K., Schafe, G. E., & Doux, J. E. Le. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(August).

Niznikiewicz, M. a., & Delgado, M. R. (2011). Two sides of the same coin: Learning via positive and negative reinforcers in the human striatum. *Developmental Cognitive Neuroscience*, *1*(4), 494–505. http://doi.org/10.1016/j.dcn.2011.07.006

O'Doherty, J., Rolls, E. T., Francis, S., Bowtell, R., & McGlone, F. (2001). Representation of pleasant and aversive taste in the human brain. *Journal of Neurophysiology*, *85*(3), 1315–21. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11248000

O'Doherty, J. P., Buchanan, T. W., Seymour, B., & Dolan, R. J. (2006). Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron*, *49*(1), 157–66. http://doi.org/10.1016/j.neuron.2005.11.014

Orsini, C. a., Trotta, R. T., Bizon, J. L., & Setlow, B. (2015). Dissociable Roles for the Basolateral Amygdala and Orbitofrontal Cortex in Decision-Making under Risk of Punishment. *Journal of Neuroscience*, *35*(4), 1368–1379. http://doi.org/10.1523/JNEUROSCI.3586-14.2015

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2010). Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(32), 10799–808. http://doi.org/10.1523/JNEUROSCI.0788-10.2010

Ramsey, J. D., Hanson, S. J., Hanson, C., Halchenko, Y. O., Poldrack, R. a, & Glymour, C. (2010). Six problems for causal inference from fMRI. *NeuroImage*, *49*(2), 1545–58. http://doi.org/10.1016/j.neuroimage.2009.08.065

Ramsey, J. D., Sanchez-Romero, R., & Glymour, C. (2014). Non-Gaussian methods and high-pass filters in the estimation of effective connections. *NeuroImage*, *84*, 986–1006. http://doi.org/10.1016/j.neuroimage.2013.09.062

Rieger, M. O., Wang, M., & Hens, T. (2014). Risk Preferences Around the World Risk Preferences Around the World, (February 2015).

Robertson, B. D., Hiebert, N. M., Seergobin, K. N., Owen, A. M., & MacDonald, P. A. (2015). Dorsal striatum mediates cognitive control, not cognitive effort per se, in decision-making: An event-related fMRI study. *NeuroImage*, *114*, 170–184. http://doi.org/10.1016/j.neuroimage.2015.03.082

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience*, *17*(11). http://doi.org/10.1038/nn.3832

Sescousse, G., Caldú, X., Segura, B., & Dreher, J.-C. (2013). Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *37*(4), 681–96. http://doi.org/10.1016/j.neubiorev.2013.02.002

Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature Neuroscience*, *8*(9), 1234–1240. http://doi.org/10.1038/nn1527

Sharot, T., Shiner, T., & Dolan, R. J. (2010). Experience and choice shape expected aversive outcomes. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(27), 9209–15. http://doi.org/10.1523/JNEUROSCI.4770-09.2010

Shenhav, A., & Buckner, R. L. (2014). Neural correlates of dueling affective reactions to win-win choices. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(30), 10978–83. http://doi.org/10.1073/pnas.1405725111

Shohamy, D. (2011). Learning and motivation in the human striatum. *Current Opinion in Neurobiology*, *21*(3), 408–414. http://doi.org/10.1016/j.conb.2011.05.009

Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2014). From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. *Neuroscience and Biobehavioral Reviews*, *46*(P1), 44–57. http://doi.org/10.1016/j.neubiorev.2013.11.003

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, *13*(8), 334–40. http://doi.org/10.1016/j.tics.2009.05.001

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. http://doi.org/10.1016/j.neuroimage.2008.03.061

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. a. (2007). The neural basis of loss aversion in decision-making under risk. *Science (New York, N.Y.)*, *315*(5811), 515–8. http://doi.org/10.1126/science.1134239

Tremblay, M., Cocker, P. J., Hosking, J. G., Zeeb, F. D., Rogers, R. D., & Winstanley, C. A. (2014). Dissociable effects of basolateral amygdala lesions on decision making biases in rats when loss or gain is emphasized, 1184–1195. http://doi.org/10.3758/s13415-014-0271-1

Tversky, a, & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science (New York, N.Y.)*, *211*(4481), 453–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7455683

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*(4). Retrieved from http://www.jstor.org/stable/10.2307/2352759

Uddin, L. Q. (2014). Salience processing and insular cortical function and dysfunction. *Nature Publishing Group*, *16*(1), 55–61. http://doi.org/10.1038/nrn3857

Wang, D. V., & Tsien, J. Z. (2011). Convergent processing of both positive and negative motivational signals by the VTA dopamine neuronal populations. *PLoS ONE*, *6*(2). http://doi.org/10.1371/journal.pone.0017047

Weber, E., Blais, A., & Betz, N. (2002). A domain□specific risk□attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, *290*(August), 263–290. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/bdm.414/full

Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S., … Johnstone, T. (2004). Human amygdala responsivity to masked fearful eye whites. *Science (New York, N.Y.)*, *306*(5704), 2061. http://doi.org/10.1126/science.1103617

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2013). Neural Evidence That Human Emotions Share Core Affective Properties. *Psychological Science*, *24*, 947–956. http://doi.org/10.1177/0956797612464242

Winston, J. S., Vlaev, I., Seymour, B., Chater, N., & Dolan, R. J. (2014). Relative valuation of pain in human orbitofrontal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(44), 14526–35. http://doi.org/10.1523/JNEUROSCI.1706-14.2014

Wolff, S. B. E., Gründemann, J., Tovote, P., Krabbe, S., Jacobson, G. a, Müller, C., … Lüthi, A. (2014). Amygdala interneuron subtypes control fear learning through disinhibition. *Nature*, *509*, 453–8. http://doi.org/10.1038/nature13258

Yokoyama, R., Nozawa, T., Sugiura, M., Yomogida, Y., Takeuchi, H., Akimoto, Y., … Kawashima, R. (2014). The neural bases underlying social risk perception in purchase decisions. *NeuroImage*, *91C*, 120–128. http://doi.org/10.1016/j.neuroimage.2014.01.036

Severe Illnesses	Severe Car Accidents	Severe Train Scenarios	Severe House Scenarios
diabetes	fender bender	bomb threat	fire
heart disease	head on collision	threatening with gun	roof collapse
lung cancer	tree falling on car	vomiting	hit by tornado
malaria	engine on fire	harassing passengers	hit by car
tuberculosis	brakes failing	threatening with knife	floor collapse
HIV/AIDs	blown tire	mugging	sink hole
brain tumor	side swiped	exposing themselves	meteor hits house
liver disease	rock through windshield	biting	carbon monoxide
Parkinson's	stuck in a ditch	hijacking train	staircase collapse
Huntington's	skid on black ice	threatening with bat	electrocution
blood poisoning	stuck in snow bank	trying to grope	fall down stairs
pneumonia	engine overheating	threatening to hit	gas leak

Appendix A. Aversive stimuli categories.

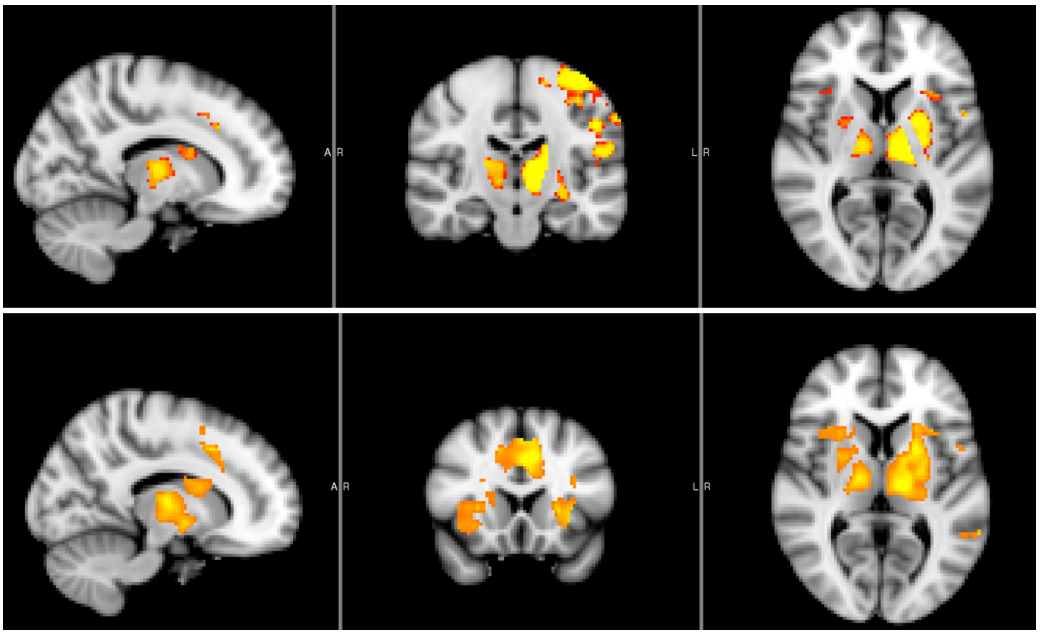
Supplemental Methods

TFCE: We used a nonparametric testing method, threshold-free cluster enhancement, to validate our results (Smith & Nichols, 2009). TFCE is a nonparametric method that calculates significant clusters of activation by estimating voxelwise cluster-like local support, which is then tested against a null distribution generated using permutation testing. Permutation testing is performed to the height of the maxima of the resulting statistic image, maintaining strong control over family-wise error. TFCE avoids the step of specifying an investigator defined threshold on clusters, which can bias results (Ecklund et al., 2016), and is sensitive to a broad range of signal shapes. The reason for using TFCE was to establish that the significant results identified testing against a null distribution generated from 5000 permutations were comparable to the results generated using a cluster correcting method, similar to the approach taken in the recent "cluster failure" follow up manuscript published in PNAS (Kessler et al., 2017).

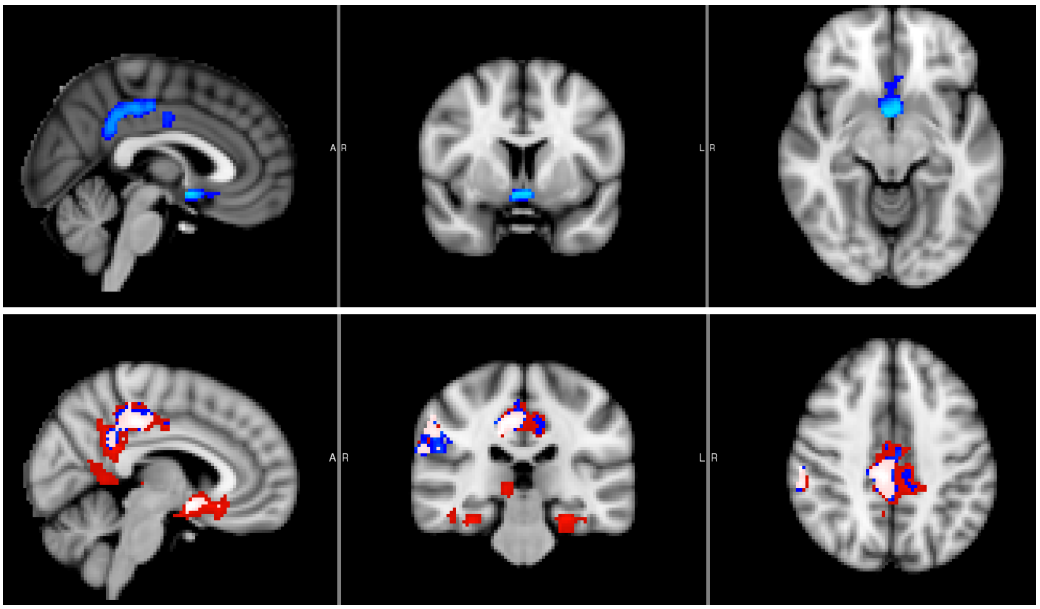
For the All Choices - Activation and All Choices -Deactivation one sample T tests reported in the Methods section, TFCE was run using 5000 permutations. Significance was defined using p=.05 corrected whole brain.

TFCE of the all aversive choices activation one-sample T test identified significant clusters of activation consistent with results produced using a cluster threshold of z=2.33, p=.05 (Supplementary Figure 1).

TFCE of the All Choices -Deactivation one-sample T test identified significant clusters of deactivation that were highly similar to the results produced using a cluster threshold of both z=2.33, p=.05 and z=3, p=.05 (Supplementary Figure 2).



Supplementary Figure 1. Top, TFCE results corrected at p=.05 whole brain of the All Aversive - Activation one-sample T test. Bottom, results of the All Aversive - Activation one sample T Test using cluster correction of z=2.33, p=.05.



Supplementary Figure 2. Top, TFCE results corrected at p=.05 whole brain for the All Aversive - Deactivation one sample T Test. Bottom, results of the All Aversive -Deactivation one sample T Test using cluster correction of z=2.33, p=.05 (red underlay); TFCE results corrected at p=.05 whole brain for the All Aversive -Deactivation one sample T Test (blue overlay); results of the All Aversive - Deactivation one sample T Test using cluster correction of z=3, p=.05 (light pink overlay).