Interactions Between Methodological and Interindividual Variability: How Monetary Incentive Delay (MID) Task Contrast Maps Vary and Impact Associations with Behavior.

Michael I. Demidenko[1], Karthikeyan Ganesan[1], Alexander S. Weigard[1,2], Hyesue Jang[1], Andrew Jahn[3], Edward D. Huntley[4], Daniel P. Keating[1,4]

1. Department of Psychology, University of Michigan.
2. Addiction Center, Department of Psychiatry, University of Michigan
3. The Functional MRI Laboratory, University of Michigan, Ann Arbor
4. Survey Research Center, Institute for Social Research, University of Michigan

Correspondence concerning this article should be addressed to Michael Demidenko, Department of Psychology, University of Michigan, 530 Church St. 2036, Ann Arbor, MI 48109. E-mail: demidenm@umich.edu

Manuscript Info. The project here was preregistered before any analytic steps were performed (https://osf.io/xh7bz). *This pre-print that has not been peer-reviewed*, however, a revised version of this preprint (under the same title) has been peer-reviewed and accepted at the journal of *Brain & Behavior* in January 2021 (per reviews published at the journal). Revisions resulted in modification to the introduction, reporting of results and interpretation in the discussion section. Please reference the supplemental materials for additional information noted in the manuscript, and see publicly shared code (https://github.com/demidenm/MIDContrasts) and the NeuroVault collection for group level statistical maps referenced in-text (https://neurovault.org/collections/6210/).

Acknowledgments. This research was supported, in part, by a grant from the *Eunice Kennedy Shriver* National Institute of Child Health & Human Development (NICHD; R01HD075806, D.P. Keating, Principal Investigator). M. Demidenko was also supported by the NICHD Developmental Psychology Training Grant (5T32HD007109-34, V.C. McLoyd & C.S. Monk). A. Weigard was supported by NIAAA T32 AA007477 (Dr. Frederic C. Blow). The authors thank Bennet Fauber, Krisanne Litinas, Christine Wagner, Hani Nasr, Peter Batra, Joshua Hatfield, Meredith House, Kyle Kwaiser, Kathleen LaDronka, the U-M Survey Research Operations staff and the U-M MRI Laboratory, for their support in collecting this data. Finally, the authors thank Leili Mortazavi and Dr. Brian Knutson for sharing code to help facilitate and plot the direction observation of BOLD signal locked to cue and phase.

Author's Contribution. MD conceived the study, MD, KG, AW conducted the statistical analysis, MD wrote the initial draft the manuscript, and KG, AW, HJ, AJ provided support with the analyses, drafts of analyses and results. DK, EH, designed and executed the survey and the

47 neuroimaging protocol. KG, AW, HJ, AJ, EH and DK assisted MD with study writing and
48 revisions. All authors read and approved the manuscript.
49 **Conflict of Interest**. The authors declare that they have no conflicts of interest.
50
51

Abstract

Phenomena related to positive emotional valence and reward responsiveness have been extensively studied in psychology. These constructs have been linked to midbrain dopaminergic pathways central to the literature on psychopathologies and development, and their measurement is a key interest in task-based fMRI. One such task, used for almost twenty years, is the Monetary Incentive Delay (MID) task. By cueing and delivering performance contingent reward, this task has been demonstrated to elicit robust and distinct activation of neural circuits involved in different phases of reward responsiveness (e.g. anticipation and outcome). Despite the broad application of the MID task, systematic evaluations of common task contrasts have been limited to between study comparisons of mean level (or group level) activation maps. In this study, we systematically examine within-task and between-contrast differences in MID task activation maps and how these differences impact inferences about their correlations with psychological characteristics. In a sample of 104 participants (Age Mean = 19.3, SD = 1.3; Female 57%), we evaluate similarities between contrasts in group- and individual-level activation maps, region-of-interest activations and their correlations with psychological characteristics. Our findings demonstrate more similarities than differences between positive and negative cues during the anticipation contrast, dissimilarity between some positive anticipation contrasts, a robust deactivation effect in the outcome phase, and behavioral associations that are less robust than previously thought. This work has practical implications for helping researchers interpret prior MID studies and make more informed *a priori* decisions about contrasts to focus on in future work. Consistent with other recent findings from large neuroimaging samples, it also suggests that researchers using the MID to identify brain-behavior relationships may have to more carefully specify their contrasts in advance in order to reliably detect small, variable effects.

Key words: Monetary Incentive Delay, Reward Processing, Approach, Avoidance, Prediction Error, fMRI, Measurement

1. Introduction

1.1. Purpose

Due to the hypothesized role of the midbrain dopaminergic reward system in wanting, liking, and learning about rewarding stimuli, neural measurements of reward processing have become a central focus in the study of various psychopathologies and problem behaviors (Berridge & Robinson, 2003; Ernst & Luciana, 2015). The Monetary Incentive Delay (MID) task, specifically, has been frequently used to measure neural correlates of approach and avoidance mechanisms during reward processing (Knutson et al., 2000). Due to its ability to probe neural circuity of reward systems, the MID task has proven a valuable indicator of dysfunction in reward related processes and various maladaptive behaviors (Balodis & Potenza, 2015; Dugré et al., 2018). More recently, the task has been incorporated into large longitudinal studies to index developmental changes in reward mechanisms and their links with negative outcomes (Casey et al., 2018; Schumann et al., 2010). Despite frequent use of this task, there are relatively few studies that have examined how analytic choices made by investigators may impact the results and interpretations about their findings in mean level activation and associations with behavior. Therefore, this study aims to understand the methodological and interindividual variability in MID task contrast maps and their associations with psychological behavior.

1.2. MID task and Central Tenets of Reward Processing

The MID task has been used in functional magnetic resonance imaging (fMRI) research for almost 20 years and is considered a robust measurement of incentive motivation (Knutson et al., 2000; Knutson & Greer, 2008). The task was designed to leverage spatial and temporal properties of subcortical structures to localize brain activation in substance use populations

(Brian Knutson & Heinz, 2015), demonstrating correlates with individual traits of positive and negative arousal (Wu et al., 2014). A central assumption of the task is that there are brain regions responsible for anticipating and responding to salient stimuli that are positively or negatively valenced, inspired in part by the literature on Pavlovian conditioning and dopamine responses to positive cues (Knutson et al., 2000). Projections from the dopamine (DA) rich ventral tegmental area (VTA) are thought to enhance activation in striatal regions that respond to reward anticipation (e.g., tones or cues that predict incentives) and in mesial prefrontal regions that respond to reward outcomes (Breiter et al., 1996; Knutson et al., 2003; Schultz, 1998). The task allows a comparison of valence (winning, positive valence, or losing, negative valence, big or small rewards) and temporal phase (anticipation or outcome). It is an *instrumental-reward task* that delivers rewards that are contingent on performance involving a timed button response (Richards et al., 2013). Different neural regions are recruited depending on whether the reward is being anticipated (or wanted) or consumed (or liked) (Haber & Knutson, 2010).

Activation patterns within these phases would be expected to align with recent theories of reward processing. For instance, the first stage during cue presentation (prior to probe, or response phase), may be modeled as a 'wanting' phase, eliciting motivation (or saliency of the reward/cue). This *anticipation phase* should elicit robust activation in striatal regions as DA has been shown to have robust effects on wanting (or incentive salience) in both animals and humans in the ventral striatum (VS) and ventral pallidum (Berridge, 2007, 2019; Berridge & Kringelbach, 2015). Conversely, when modeling the *outcome phase* (or liking), one would expect less activation of VS (as only ~10% of neurons in NAcc facilitate pleasure) in response to the pleasure of reward, as hedonic 'hot-spots' are more likely to be represented in the insula and OFC (Berridge & Kringelbach, 2015).

The bulk of fMRI analyses using the MID task have focused on specific, unmodulated phases of the task. But, previous work suggests that modulators based on formal models of reinforcement learning may be important to incorporate into the task to account for individual variability not captured by a standard analysis (Bjork et al., 2010; Oldham et al., 2018). Although the utility of prediction error is still debated (Berridge & O'Doherty, 2014), it remains to be seen whether or not, as part of a prediction error model, expected value and prediction error (positive or negative) modulates the signal in the anticipation and outcome phases in the MID task. Such modulators may be critical in accounting for individual level variation that drives performance and learning values that may be represented in subcortical and cortical neural signatures (Balleine & O'Doherty, 2010). Although previous work has recommended the use of modulators in the MID task (Bjork et al., 2010; Oldham et al., 2018), to our knowledge, modulators of prediction error are still underexplored. One recent study (Cao et al., 2019) using prediction error modulators found that prediction error was positively related to activation in the bilateral VS, and another found association with substance use problems in young adults (Cao et al., 2020).

1.3. Differential use and Research Degrees of Freedom in MID Task

Although the MID task has been used extensively to study dysfunctional reward processing in populations with substance use disorders (Balodis & Potenza, 2015), it has also been incorporated into other studies of neurodevelopment and broader psychopathology. Various versions of the MID task have been used to investigate reward related changes as a function of age (Bjork et al., 2010; Dhingra et al., 2019), social vs non-social rewards (Schwartz et al., 2019), psychosocial characteristics of impulsivity and sensation seeking (Büchel et al., 2017; Cao et al., 2019; Joseph et al., 2016), early adversity (Boecker et al., 2014; Gonzalez et al., 2016), substance use (Aloi et al., 2019; Cope et al., 2019; Nestor et al., 2019; Sauder et al., 2016;

Swartz et al., 2019), depression (Chan et al., 2016; Colich et al., 2017; Landes et al., 2018; Mori et al., 2016) and other clinical related problems (Bourque et al., 2017; Lancaster et al., 2016; Maresh et al., 2019; Mikita et al., 2016; Papanastasiou et al., 2018; Stevens et al., 2018; Urošević et al., 2016; Veroude et al., 2016; von Rhein et al., 2015; Xu et al., 2017). Similar to earlier reviews (Balodis & Potenza, 2015; Oldham et al., 2018), the research studies cited above often used different MID versions and incorporated various contrasts to derive activation maps that were used to compare their variables of interest (see Supplementary Table S2). This raises the question: To what extent are similarities and differences between the findings and conclusions from these studies due to variance in analytic methods?

Recent evidence has suggested that analytic methods and decisions may not only alter outcomes, but also result in different interpretations of fMRI analyses. For instance, Hong et al. (2019) found that peak level coordinates from various studies have a high degree of variability, that may often lead to inaccurate conclusions about replication. Although activations may be close in distance between two groups (or studies), such that they appear to be in similar brain regions, these may not be related to a 'replication' of a neural process that is hypothesized, due to a lack of neural specificity. In addition, Botvinik-Nezer and colleageus (2020) had 70 different teams analyze identical fMRI data, testing pre-defined hypotheses associated with risky decision-making. As a function of between lab differences, they found that both lab workflows and interpretations by researchers altered findings, even though statistical maps may have been comparable. Thus, such differences in workflow, contrast and parameter selection, and outcomes investigated are important to consider when forming substantive interpretations of fMRI findings. Without a clear understanding of how analytic decisions impact our results and

interpretations, the flexibility of fMRI analyses (e.g., "researcher's degrees of freedom") may result in an unacceptable number of false positives (Gelman & Loken, 2014).

In the MID task, for example, it is not well understood how investigators' analytic choices, such as choice of contrast (for example, Big Win Cue ($5) vs Neutral Cue, or Win Cues ($5 & $0.20) vs Neutral Cue), may impact their inferences about the relations between neural reward circuitry and behavior. FMRI activation maps differ as a function of reward type/magnitude (Bjork et al., 2010) and recent reviews suggest there is substantial variability across studies in the techniques used to derive such maps (Balodis & Potenza, 2015; Dugré et al., 2018; Oldham et al., 2018). Balodis & Potenza (2015) attempted to reconcile activation differences in addictive behaviors as a function of analytic strategies and individual level effects, but it was unclear whether these differences were related to sample characteristics or true between contrast task differences, some of which may go unreported. Combined with the file drawer problem, the diverse sets of analyses may contribute to underreported contrasts and associations with behavior which may not fit into a latent distribution of results (Gelman & Loken, 2014; Simmons et al., 2011). Therefore, it is important to quantify within-individual variation across activation maps within the same sample and assess the relative utility of these maps for predicting behavioral outcomes. This would demonstrate whether there is a) stability within individuals' estimates of activation at each phase of the task (anticipation or outcome); b) consistency between contrasts in the level of activation in specific regions of interest (ROI); and c) whether contrast choice alters the utility of these activations for predicting various psychological characteristics.

1.4. Current Study

Previous reviews of the MID task have evaluated general utilization of the task in studies of reward responsiveness (Lutz & Widmer, 2014), between-study, temporal and phase-related differences in MID activation effects (Oldham et al., 2018), dynamics of reward versus loss (Dugre 2018), and influences of substance use (Balodis & Potenza, 2015) and psychosis (Radua et al., 2015) profiles on activation differences. However, the extent to which contrast choice contributes to within-subject variability in activation maps and alters conclusions about associations between neural responses and behavior is still unclear. The current study leverages a community sample of late adolescents/emerging adults to examine variability across various activation maps in the MID task.

In order to delineate variability across contrast types (which is difficult to evaluate between samples/studies), we perform multiple common analyses that focus on the anticipation, outcome, and prediction error parameters, with data from the same individuals. Due the predominant role of motivation (or anticipation of reward) in this task, and difficulty to temporally differentiate the outcome phase (Bjork et al., 2010), we predominantly look at the anticipation phase. These activation maps are thresholded to compare the degree to which statistical maps (from ten contrasts) a) vary within a phase (for example, anticipation Big Win > Neutral contrast) and b) vary between phases of the task (for example, anticipation vs outcome). The degree of variability is assessed at the individual and group level to assess whether contrast activation patterns differ across individual subjects' results and group-level results. Then, mean signal intensity values for key regions from previous reviews, such as the insula, mPFC, OFC, VS, and amygdala (Balodis & Potenza, 2015; Dugré et al., 2018; Oldham et al., 2018) are extracted to evaluate whether activation in these ROIs from different contrasts indexes similar or divergent individual

difference dimensions. Due to recent concerns that some multiband sequences may alter the BOLD signal in subcortical regions (Risk et al., 2018), we complement these results with signal-to-noise ratios and plotted time-series from the VS to provide a direct observation of signal for each anticipation condition. Finally, correlations between these ROI activations and self-reported psychological characteristics (aggregate score of multi-wave longitudinal data) are assessed to determine the impact of contrast choice on the prediction of psychological measures including substance use, psychosocial, and socioemotional functioning.

Similarities and differences are not intended to be reported within a null hypotheses test framework, but rather presented as a statistical index of overlap (Jaccard's similarity coefficient) and of associations across ROIs and behavior (Pearson's *r* coefficient; heat maps of *r* point estimates for inter-ROI relationships and posterior distributions of *r* values for associations of ROIs with behavioral covariates). Our broad goal is to improve the field's understanding of how and where there is within-task variability as a function of MID task contrast choice, and, in doing so, to inform the interpretation of existing MID studies and better guide researchers' *a priori* decisions about contrasts to focus on in future studies. Due to the exploratory nature of the analyses, the background, methods and analytic plan were preregistered on the Open Science Framework (https://osf.io/xh7bz). However , we elected not to preregister specific hypotheses because the intended purpose of the study was to use exploratory analyses to provide a holistic overview of how degrees of freedoms impact interpretation of MID task results (Thompson et al., 2020).

2. Methods

Participants in this neuroimaging study are from a subsample of the Adolescent Health Risk Behavior (AHRB) study. AHRB consists of a nonprobability sample of 2017 (mean age =

16.8, *SD* = 1.1; Female 56%) 10th and 12th grade students recruited from nine public school districts across eight Southeastern Michigan counties, using a quota sampling method to enhance sample diversity. Phase I, described elsewhere (Demidenko et al., 2019), collected demographic, psychosocial, neurocognitive and behavioral information across three waves. From Phase I of the study, a subsample of 115 adolescents was recruited to participate in the neuroimaging phase of the study (elapsed time between Wave 1 and neuroimaging section (Months): *M* = 30.9 months *SD* = 5.0 months). Of the 115 adolescents that participated, 108 completed the magnetic resonance imaging (MRI) portion of the visit. Seven participants were ineligible or unable to participate in the MRI due to not meeting MRI safety eligibility (e.g. claustrophobia [*n* = 3] or no formally documented medical clearance to rule out potential metal in body [*n* = 4]). Of the 108 participants that completed the MRI, four participants were excluded from the analyses due to: artifacts in the images that were not recoverable, and one participant that stopped responding during the second run of the task. The final fMRI subsample (N = 104; Age Mean = 19.3, *SD* = 1.3; Female 57%) was included in the subsequent analyses and did not differ from the full sample in age, gender, or time from the original survey. The bulk of code used in the subsequent analyses are made available online (https://github.com/demidenm/MIDContrasts).

2.1. Self-Reported Psychological Measures

Substance Use. Substance use behaviors (marijuana and alcohol) are assessed via the item: "On how many occasions (if any) have you [used marijuana or hashish/had any alcoholic beverage to drink—more than just a few sips] during the last 12 months?" Responses are reported on a seven-point scale ranging from 1 = "0 occasions" to 7 = "40 or more occasions". Use is further probed for past 30-day use for alcohol and marijuana. For alcohol, last 30-day use assessed binge drinking occasions, "During the last 30-days, how

many times (if any) have you had four (for females)/five (for males) or more drinks in a row, that is, within about 2 h?" Response options were, none, once, twice, 3 to 5 times, 6 to 9 times, or 10 or more times. For marijuana, use occasions within the last 30-days were assessed using the question, "How many times (if any) have you used marijuana or hashish during the last 30-days?" Response options were reported on a seven-point scale ranging from 1 = '0 occasions' to 7 = "40 or more occasions". Substance use items are identical to those used in the annual, national Monitoring the Future surveys (Johnston et al., 2018). Marijuana and alcohol scores were z-scored, and then a substance use aggregate measure was created by averaging the z-scored items across Wave 1 – Wave 3.

Impulsivity. The Barratt Impulsiveness Scale-Brief (BIS-B) is an 8-item, unidimensional measure of impulsiveness (Steinberg et al., 2013) based on a reduced item set obtained from the Barratt Impulsiveness Scale (BIS), 11th revision. Items were rated on a 4-point Likert-type scale: rarely/never (1), occasionally (2), often (3), and almost always/always (4). A mean score was computed (range: 1 – 4), higher scores indicated lower self-reported impulsivity ($\alpha = .79$). An aggregate score was created by averaging scores across Wave 1 – Wave 3.

Sensation Seeking. The Brief Sensation Seeking Scale (BSSS) is an 8-item self-report measure of sensation seeking (Hoyle et al., 2002) based on reduced item set of the Zuckerman Sensation Seeking Scale (SSS). The items measure dimensions of sensation seeking: experience seeking, boredom susceptibility, thrill and adventure seeking, and disinhibition. Responses were on a 5-point Likert-scale: strongly disagree (1), disagree (2), neither disagree nor agree (3), agree (4), and strongly agree (5). A mean score was computed (range: 1–5), with higher scores indicated higher self-reported sensation seeking ($\alpha = .78$). An aggregate score was created by averaging scores across Wave 1 – Wave 3.

Socioemotional problems. Socioemotional problems were assessed using the Youth Self-Report (YSR; (Achenbach & Rescorla, 2001) to characterize externalizing and internalizing problems. The YSR is a widely utilized, 112-item self-report measure assessing emotional and behavioral difficulties in 11-18-year-olds. The YSR includes two broadband scales: internalizing problems (e.g. withdrawn/depressed) and externalizing problems (e.g. attentional deficit/hyper activity problems, oppositional defiant problems). Raw scores are normalized to provide a common metric with higher scores indicating greater psychopathology. Validity and reliability of the YSR broadband, syndrome, and DSM-oriented scales are well documented (Achenbach, 2013; Achenbach & Rescorla, 2001) with adequate internal consistency (α = .70 - .86) and test-retest reliability (α = .67 - .88). An aggregate score was created from population-standardized t-scores for internalizing and externalizing by averaging scores across Wave 1 – Wave 3. In the present study, Cronbach's alphas of .91 and .88 were obtained for the internalizing and externalizing scales, respectively.

2.2. fMRI Task

A modified version of MID task (Knutson et al., 2000) was used to model neural signatures of the anticipation and outcome of monetary rewards (Bjork et al., 2010; Büchel et al., 2017; Cao et al., 2019). The MID (Knutson & Greer, 2008) is an established task for assessing reward processing, and the modified version in this study (a modification from the Michigan Longitudinal Study, Martz et al., 2016) is currently being employed in the national Adolescent Brain Cognitive Development (ABCD) study to measure the development of adolescent reward processing (Casey et al., 2018). Identical to the task described in Casey et al. (2018), the task in this study consists of three phases: anticipation, probe and outcome (that is, feedback). Each trial starts with a cue type (Win $0.20, Win $5, Lose $5, Lose $0.20, or No Money At Stake). There

are twelve trial orders of the task, consisting of 50 contiguous trials and 10 trial types per run (5:42 minutes long). Participants completed two runs of the MID task during the scan (100 trials and 20 trial types). The task is individualized to reach around 60% accuracy rate by adjusting the difficulty (that is, probe duration). **See Section 1.1 in Supplementary Materials** for more information on task paradigm and administration. A key difference between the current version of the MID and that used in the IMAGEN sample (Cao et al., 2019), is the latter only includes Win and neutral trials, excluding Loss trials.

2.3. fMRI Data Acquisition and Preprocessing

Data were acquired using a GE Discovery MR750 3.0 Tesla scanner with a standard adult-sized coil (Milwaukee, WI). A full-brain high-resolution T1 SPGR PROMO scan was acquired that is used in preprocessing (TR = 7000ms, TE = 2900ms, flip angle = 8°, FOV = 25.6 cm, slice thickness = 1 mm, 208 sagittal slices; matrix = 256 x 256). Before the MID task, a fieldmap was acquired using spin-echo EPI (TR = 7400ms, TE = 80 ms, FOV = 21.6 cm, 90x90 matrix) with opposite phase encoding polarity (A→P, P→A). Two functional T2*-weighted BOLD MID runs were acquired in the axial plane following structural and a faces task using a multiband EPI sequence (MB factor=6) of 60 contiguous axial 2.4 mm slices (TR = 800ms, TE = 30 ms, flip angle = 52°, FOV = 21.6 cm, 90x90 matrix, volumes = 407).

fMRI Data Analyses

FMRI data were reconstructed and realigned using SPM12, physiological noise was removed using RETROICOR, and a fieldmap correction was applied in SPM12 to each T2* run to recover inhomogeneity of signal in the B0 field. Preprocessing steps were completed using FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) FEAT (FMRI Expert Analysis Tool) Version 6.00. After volumes were (1) reconstructed, (2) realigned, (3) physiological noise was

328 removed and (4) field map correction was applied, the following preprocessing steps were
329 performed: (6) registration to high resolution structural and standard space MNI 152 image using
330 FLIRT using a Full search 12 DOF (Jenkinson & Smith, 2001; Jenkinson, Bannister, Brady, &
331 Smith, 2002), (6) motion correction using MCFLIRT (Jenkinson et al., 2002), (7) non-brain
332 removal using BET (Smith, 2002), (8) spatial smoothing using a Gaussian kernel of FWHM
333 5mm, (9) grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative
334 factor and (10) highpass temporal filtering (Gaussian-weighted least-squares straight line fitting,
335 with sigma=50.0s).

336 3. fMRI Analyses

337 Subjects were excluded from analyses if a subject's mean framewise displacement (FD)
338 values exceeded > .9 within any given run (Mean FD Pre- & Post-preprocessing included in
339 **Supplementary Section 1.2**), all subjects mean post FD were < .9. We focused on commonly
340 used contrasts (Table 1) from a recent review (Oldham et al., 2018) and those from our review of
341 studies using the MID (PubMed 2015 – 2019; **Supplementary Table S2**), such as reward
342 anticipation (such as Big Win or Lose ($5), Small Win or Lose ($0.20) versus neutral

Table 1: Contrast Modeled in the Monetary Incentive Delay Task

Contrasts	Phases of MID Modeled
Contrast 1 (A1) - Ant	Win (W; $5 & $0.20) > Neutral (N) (W>N)
Contrast 2 (A2) - Ant	Big Win (BW; $5) > Neutral (N) (BW>N)
Contrast 3 (A3) - Ant	Big Win (BW; $5) Small Win (SW; ($0.20) (BW>SW)
Contrast 4 (A4) - Ant	Big Win (BW; $5) > Implicit Baseline (BW>IB)
Contrast 5 (A5) - Ant	Big Loss (BL; $5) > Neutral (N) (BL>N)
Contrast 6 (F6) – Out	Big Win (BW; $5) Hit > Neutral (N) Hit (BWH>NH)
Contrast 7 (F7) – Out	Big Loss (BW; $5) Hit > Neutral (N) Hit (BWH>NH)
Contrast 8 (P8) - PE	Expected Value – BW & SM Modulated (EV)
Contrast 9 (P9) - PE	Positive Prediction Error (PE) - BW & SM Modulated (PPE)
Contrast 10 (P10) - PE	Negative Prediction Error (PE) - BL & SL Modulated (NPE)

Ant = Anticipation; Out = Outcome; Individual contrasts modeled in FSL, see section 1.4 in Supplementary for list of Events Modeled in GLM. A = Anticipation; F = Feedback; P = Prediction Error

anticipation, win outcome (such as $5 or $0.20) versus neutral outcome, loss conditions (such as $5 or $0.20) and alternative contrasts that may be comparable to test for similarities within a group, for example, gain or big gain conditions. It should be noted that using anticipation vs outcome phase yields in estimates that are often powered differently, as a function of the target accuracy of the task (60%) leading to individual variation in hit/miss trials. Furthermore, since the outcome phase is often difficult to deconvolve in the task, and modeled in various ways (see **Supplementary Table S2)**, we include one type of outcome contrast focusing on gain and loss, as it is not a central focus of these analyses and often not the focus in contrasts in the literature.

First-level analyses were performed by using FEAT. Time-series statistical analysis was carried out using FILM with local autocorrelation correction (Woolrich et al., 2001). Similar to other studies (Cao et al., 2019; Hagler et al., 2019; Lamm et al., 2014), both Anticipation and Outcome events were modeled (15 explanatory variables) and modulated prediction error signal of EV, PPE and NPE (see Table 1), in addition to six motion parameters (translations and rotations in x, y, z directions) and the derivatives of the motion parameters. The modeled contrasts and design matrix are described in greater detail in **Supplementary Section 1.3**. We included prediction error explanatory variables based on a recent review suggesting the MID is considered to be an implicit reinforcement learning (RL) paradigm (Balodis & Potenza, 2015), and others recommending use of modulators (Bjork et al., 2010; Oldham et al., 2018). To incorporate these recommendations, the RL modulators included: Expected Value (EV) and Prediction Error (PE). To derive estimates of EV and PE for this task, the behavioral data were modeled for each participant (100 trials – trial-by-trial) to calculate parametric modulators (EV for anticipation; PE for Received Reward (RR); *pGain* = probability gain, η = learning rate

(0.7)). Similar to Cao et al. (2019), we used a RL model trained by reward cues and outcomes (Rescorla & Wagner, 1972):

$$EV_t = pGain_t \times Cue_t$$

$$PE_t = RR_t \times EV_t$$

$$pGain_{t+1} = pGain_t + \left(\eta \times \frac{PE_t}{Cue_t}\right)$$

To average across the two runs that are used in subsequent stages, a second-level model was defined for each participant for each of the ten contrasts (**see Supplementary Section 1.3)** using fixed effect analysis in FEAT. A group-level analyses was performed using FMRIB's Local Analysis of Mixed Effects (FLAME 1) to generate a mean level activation across subjects for a given contrast. Considering the large array of contrasts that are modeled, abbreviations from the first column of Table 1 are referred to when referencing contrasts henceforth.

To provide a direct observation of the BOLD signal and signal-to-noise information of subcortical regions, we include complementary post-hoc analyses evaluating raw BOLD signal (see Section 2.6 in Supplemental Materials). We extract the mean signal for VS and mPFC in the timeseries for VS and plot it for 15 TRs. Likewise, for cortical mPFC and subcortical VS we extract and present the distribution of the signal-to-noise ratios (SNR) for each individual and run to confirm that SNR is within an acceptable range (see Section 2.6 in Supplemental Materials).

3.1. Individual Level and Group Estimates

In order to compare overlap between thresholded activation maps for each contrast at the individual and group level, we thresholded activation maps produced by the second level and group level analyses. For the individual level, subjects second level maps (zstat) for each contrast are thresholded at $p < .01$ ($z = 2.3$) and group level contrasts are thresholded at $p < .001$ ($z = 3.1$).

We selected a lower threshold for individual maps due to more variability in estimates within an individual map, that may substantially alter the Jaccard's Similarity Indices. These thresholded maps are binarized (using *fsl* -bin) and compared to derive Jaccard's Similarity Indices (described below).

3.2. Calculating Similarity

One of the aims for this study is to compare similarity between different activation maps of the MID task within individuals and at the group level. This requires calculating similarity matrices between our ten contrasts (described above). FMRI task activation maps have been utilized in similar studies before (Maitra, 2010), but there is no consensus on how to measure percent overlap between thresholded activation maps (McGonigle et al., 2000). While measures such as Dice coefficient (Dice, 1945) have been widely used to calculate percent overlap (Taha & Hanbury, 2015), it possesses undesirable properties (Tulloss, 1997).

This led us to choose Jaccard similarity index (JSI) to calculate percent overlap. One of the major advantages of using the Jaccard similarity index is that the percent overlap results obtained from this technique are intuitive and physically interpretable (Maitra, 2010). The percent overlap between any two activation maps is defined from a set theoretical point of view, where the overlap $J(A, B)$ is defined by the well-known relation as:

$$J(A,B) = \frac{A \cap B}{A + B - A \cap B}$$

The relation calculates the ratio of common pixels that are activated across two activation maps to the total number of pixels present in the two maps. Experimental results that utilized this metric in previous fMRI reproducibility studies can be found in Maitra (2010).

With the JSI as our point estimates in evaluating replicable results across different contrasts, we propose a bootstrapping based confidence interval calculation for identifying the

95% confidence intervals of the overlap measures across all subjects in our sample. Bootstrap-based approaches have been popular in calculating robust approximate confidence intervals (DiCiccio & Efron, 1996). We propose this approach for identifying the bounds of percent overlap between activation maps across subjects for two reasons: 1) It would provide reliable estimates of the range and shape of the distribution of percent overlap, 2) It would provide a physical interpretation of the JSI obtained across all of the subjects.

3.3. Region of Interest and Behavioral Associations

Voxel coordinates, from Neurosynth.org, for a priori ROI's, bilateral insula, OFC, VS, and mPFC and ACC (see supplemental Table S1 and Figure S1), were used with *fslmaths* to create 10mm-diameter spheres. For each ROI, the voxels from each contrast mask (using z-statistics produced by Feat Second Level) are averaged to create a mean signal intensity value and were extracted using *fslmeants.* Correlations (point estimates of Pearson's *r*) across ROIs were analyzed in R version 3.6.1 (R Core Team, 2019) and were visualized using a heatmap.

ROI mean level signal intensity values across ten contrast types (described above), were used to assess associations between neural activity and self-reported aggregate scores of a) substance use, b) sensation seeking, c) impulsivity, d) externalizing, and e) internalizing problems. Bayesian correlation analyses implemented in JASP (JASP Team, 2019; Ly et al., 2018) were used to estimate posterior distributions for the Pearson's *r* value of each predictive association. Default, non-informative priors (uniform distributions spanning the values from -1 to 1) were used for all correlation analyses. Median values of the posterior distribution, which indicate the most likely *r* value, and 95% credible intervals, which represent the lower and upper bounds of the range which has a .95 probability of containing the *r* value, are reported below to quantify the strength of, and uncertainty about, these predictive associations. As analyses are not

intended to be formal tests of hypotheses, we will refrain from reporting either Bayes factors or frequentist *p*-values.

4. Results

4.1. Demographics, Task Behavior and General Overview

The demographic characteristics for the full sample (n = 104) are provided in Supplementary Section 2.2**,** Table S3. For the anticipation phase (A1-A5) and prediction error models (P8-P10), all 104 individuals were included, however, for the feedback phase (F6 & F7) four subjects were excluded due to underpowered conditions resulting in anomalies in the estimated [First & Second Level] statistical maps, resulting in only N = 100 for those contrasts. The behavioral performance statistics from the MID task are included in Supplementary Section 2.3**,** Table S4 and Figure S2. Although the average accuracy for the task, 57%, was below the targeted 60%, the Big Win ($5) and Big Loss ($5) conditions were at or above the target, 62% and 60% accuracy, respectively. As expected, accuracy was lower (48%) and more variable during the neutral condition. Mean response times are not reported, as the E-Prime data wasn't collected for incorrect ('miss') trials during the MID task.

Similarity matrices and activation maps are displayed in supplementary Figure S4 and Figure 1, respectively. Associations between individual differences in ROI activation estimates from each contrast are reported at https://osf.io/a5wem/, and are selectively reported below for clarity (Figure 2). Correlations between ROI activation estimates and behavioral criterion measures are reported in Figure 3. There were three notable patterns present in these results (Note: we remind the reader to refer to Table 1 for contrast descriptions): 1) during the anticipation phase, **A2:** BW>N and **A5:** BL>N demonstrate comparable striatal/insula activation and task-negative deactivation; 2) during the outcome phase, **F6:** BWH>NH and **F7:** BLH>NH

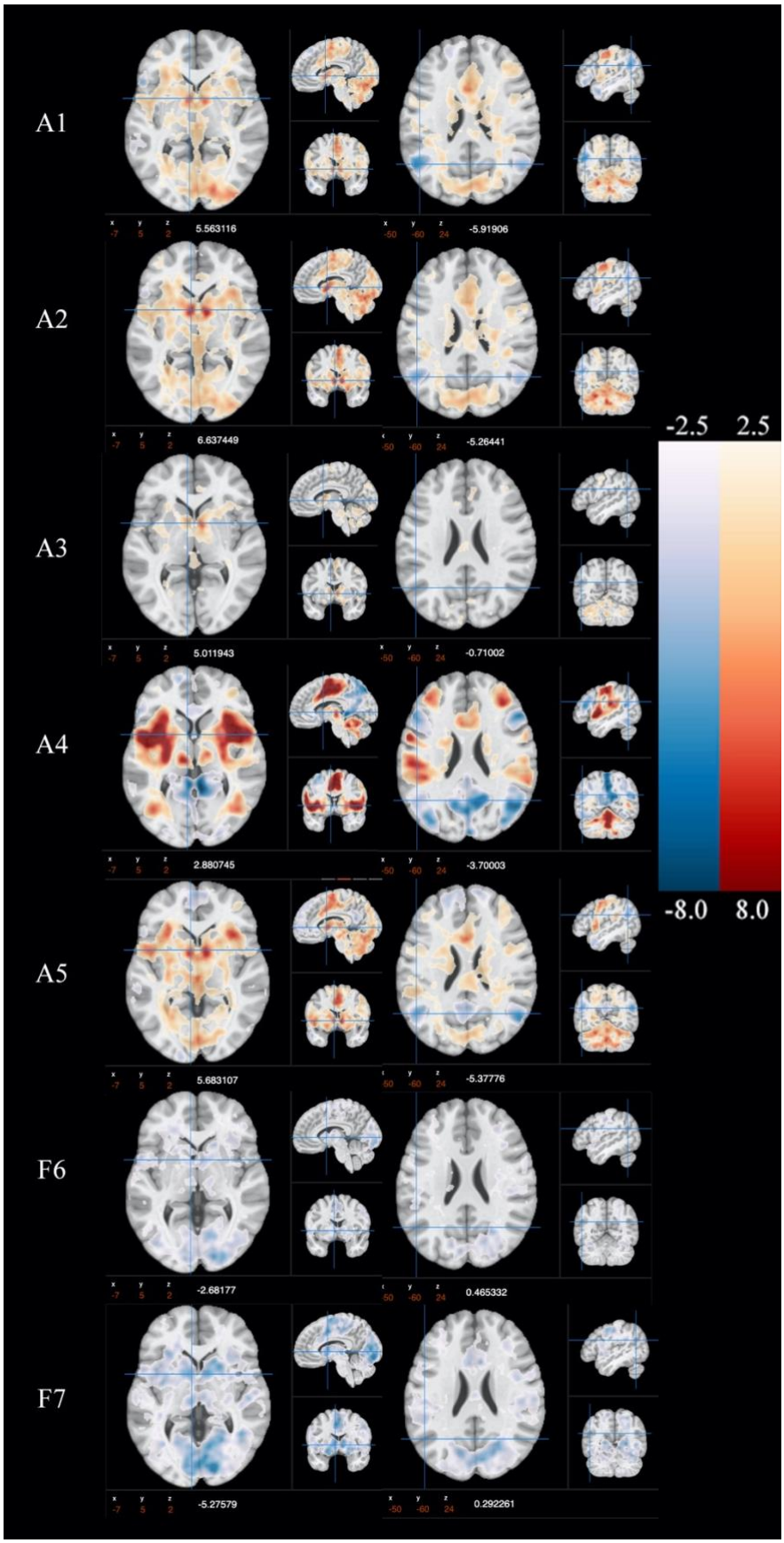

Figure 1. Mean level activation and deactivation maps for A1-A5 & F6-F7, One-Sample T-test
See Table 1 for details and online collection for untresholded statistical maps of tens contrasts
http://neurovault.org/collections/JVXLTPHC

demonstrate comparable deactivation of striatal regions; 3) **F7:** BWH>NH appears less meaningful, with reduced salience response and negligible task-negative activation and less association with other contrasts of anticipation phase; and 4) individual differences in ROI activation, across different contrasts, demonstrate relatively weak associations with behavior. The aforementioned are expanded in greater detail below. Notably, the activation maps of the prediction error models **P8:** EV, **P9:** PPE, and **P10:** NPE, were extremely variable in activation and relatively weak in their associations with mean ROI activation from other contrasts, they

are not discussed below. The contrast maps are available online.

4.2. Big Win and Loss Anticipation Engage Similar Neural Systems

The thresholded masks ($p < .001$) of **A2:**BW>N and **A5:**LB>N group maps had a Jaccard's similarity Coefficient of .16 (supplemental Figure S4). This similarity is also apparent in the group level activation maps (Figure 1), demonstrated by shared patterns of activation. Bjork et al. (2010) reported the greatest striatal activation in the largest magnitude of reward, e.g. $5 (**A2:**BW>N). Although the peak left striatal activation in the **A2:**BW>N is greater than in the **A5:**BL>N (based on magnitude of *z*-statistic in activation maps), in their direct comparison (https://neurovault.org/collections/JVXLTPHC/images/359857/), the difference is relatively small. The greatest difference between these two contrasts was increased activation in the mPFC in **A2:**BW>N as compared to **A5:**LB>N. Furthermore, contrasts **A2:**BW>N & **A5:**BL>N show similar activation of supplementary motor area (SMA), the insular cortex, thalamus and cerebellar regions. Similar to the shared positive activation of these contrasts, they, too, share comparable deactivation in the task-negative, angular gyrus, an effect that is not seen in the **A3:**BW>SM (Figure 1). This activation in the striatal regions and deactivation in task negative regions is comparable to a recent meta-analysis (open source activation maps: https://neurovault.org/collections/4258/) showing similar robust patterns of activation and deactivation in both win and loss anticipation (Wilson et al., 2018).

Consistent with these similarity analyses in group level activation, correlations of mean signal intensity values from ROIs across **A2:**BW>N and **A5:**BL>N (Figure 2, full matrix available at https://osf.io/a5wem/) also suggested that neural responses from these contrasts index similar individual difference dimensions. Positive correlations in neural responses between the contrasts were identified (Figure 2) in the anterior cingulate cortex (ACC; $r = .58$), medial

prefrontal cortex (mPFC; r = .26), bilateral Insula (Right: r = .58; Left: r = .24), bilateral orbitofrontal cortex (OFC; Right: r = .43, Left: r = 27), and bilateral ventral striatum (VS; Right: r = .57, Left: r = .42). The similarity between **A2**:BW>N and **A5**:BL>N is great than the past literature has eluded to.

4.3. Reward and Loss Outcome is Paradoxically Linked to Striatal Deactivation

Contrary to past work focused on striatal activation during win conditions, our contrasts during outcome phase, **F6:**BWH>NH & **F7:**BLH>NH, demonstrated a *deactivation* of the striatal regions. Based on the Jaccard's similarity Coefficient, .34, the regions that were deactivated were comparable in **F6:**BWH>NH and **F7:**BLH>NH (Figure 1, and supplemental Figure S4). Although the mean level deactivation of the striatal region in the **F6:**BWH>NH contrast was relatively weak (t = -2.68), in the **F7**:BLH>NH condition the deactivation was relatively robust (t = -5.8). As a control comparison in change of activation, we reference the angular gyrus, which has a relatively weak mean level activation in both **F6:**BWH>NH and **F7:**BLH>NH, demonstrating that there is a more profound change in activation in the striatal region between the anticipation and outcome phase (see Figure 1). In a direct comparison of **F6:**BWH>NH & **F7:**BLH>NH (https://neurovault.org/collections/JVXLTPHC/images/359858/), **F6:**BWH>NH demonstrates greater activation in the left parahippocampal (z = 4.3) and right nucleus accumbens (z = 3.4). These two outcome contrasts demonstrated some associations (Figure 2) in individual differences analyses of mean signal intensity in the ACC (r = .33), mPFC (r =.55), and bilateral VS (r = .45 - .46) (full matrix available at https://osf.io/a5wem/ However, this effect may relate to the fact that Big Win/Lose Hit versus Neutral Hit rather than an alternative contrast that is used in the litearture, such as Big Win/Lose Hit versus Big Win/Lose Hit versus Big Win/Lose Miss. For this latter contrast, we plot the underlying BOLD

signal in the latter contrast, and find no differentiation in striatal regions (Figure S7). Overall, there a consistent deactivation in these regions after the peak rise during the anticipation phase.

4.4 Anticipation Big Win versus Small Win Contrast Is Distinct from other Anticipation Contrasts

Despite its variable use in the literature, **A3:**BW>SM was unique when compared to other contrasts in anticipation phase (Figure 1). The **A3:**BW>SM had lowest Jaccard Coefficient's with other contrasts modeling the anticipation phase, <.02 (Figure S4). Further, in the group mean-level activation, compared to **A1:**W>N, **A2:**BW>N, and **A5**:BL>N anticipation phases, the **A3:**BW>SM had the weakest mean-level striatal and insular activation, and no task-negative activation. The task-negative activation difference is unique, as all of the other contrasts demonstrate this profile of task-negative activation in the anticipation phase.

However, with respect to individual differences in ROI mean-level activation, depending on the contrast, there are similarities between **A3:**BW>SM and other contrasts (Full correlation matrix available at https://osf.io/a5wem/). For example, the mean-level activation between **A1:**W>N and **A3:**BW>SM is negligible: ACC (r = .15), mPFC (r = -.04), bilateral insula (r = .07 - .12), bilateral OFC (r = -.14 - .04) and bilateral VS (r = .05 - .11). Yet, there is a strong association between **A2**:BW>N and **A3:**BW>SM in the ACC (r = .63), mPFC (r = .65), bilateral insula (r = .42 - .64), Right OFC (r = .62), and bilateral VS (r = .43 - .65). Despite the similarity discussed between **A2:**BW>N and **A5:**BL>N above, there is a negligible association between ROI's in **A3:**BW>SM and **A5:**BL>N (r = -.11 - .19). Which suggests that the similarities between **A2:**BW>N and **A3:**BW>SM may arise from the shared Big Win cue in the subtraction.

4.5 Across Contrasts, Activations Show Only Weak to Negligible Correlational Relationships with Behavioral Criterion Measures

The aggregated scores for psychological characteristics in this sample were associated in the direction expected (Supplementary Section 2.4, Table S5). More specifically, there was a strong positive association between internalizing and externalizing problems (r = .51), sensation seeking and impulsivity (r = .44), externalizing and substance use (r = .51), and substance use and sensation seeking (r = .38) and impulsivity (r = .24).

Posterior medians and 95% credible intervals (CIs) of Pearson's r values, which represent the most likely r value and range in which there is a .95 probability that the r value falls, respectively, are displayed in Figure 3 for all relationships between ROI activation estimates and behavioral criterion measures (complete values available at https://osf.io/d9k3v/, bootstrapped values also provided which are comparable at https://osf.io/dr5y2/). Although the interpretation of individual associations is complicated by the large number of tests reported, several general patterns are apparent.

First, the most likely r values for the majority of associations fell at or well below the threshold for what is typically considered a "small-sized" effect ($|r|$ = .10). Similarly, the bulk of most CIs also fell in this general range. In fact, there was not a single association for which the most likely r value indicated a "moderately-sized" effect ($|r|$ >= .30), and few CIs overlapped with this "moderate" criterion. It is also notable that only a handful of CIs (less than 5%) did not overlap with 0, suggesting that even these cases, which might be interpreted as showing promising evidence for a non-negligible effect, are likely due to multiple testing rather than reflecting true relationships. Indeed, as typical Bayesian CIs do not take into account the probability that the null (r = 0) is true (van den Bergh et al., 2019), the effect size estimates we report are, if anything, likely to be overly optimistic. Hence, consistent with other emerging findings from large, diverse neuroimaging data sets (Nees et al., 2012; Paulus et al., 2019; Paulus

& Thompson, 2019), these patterns of results suggest that direct associations of MID task activations with relevant behavioral criterion measures are less robust than what has been previously thought, and that even if these associations exist, effect sizes are likely to be small.

4.6 Post-Hoc Analyses

In light of prior meta-analytical comparisons of base contrasts within individuals, such as gain versus outcome phases (Knutson & Greer, 2008; Wilson et al., 2018), we compared these differences in the anticipation phase, **A2:**BW>N versus **A5:**BL>N; outcome phase, **F6:**BWH>NH versus **F7:**BLH>NH; win anticipation versus win outcome, **A2:**BW>N versus **F6:**BWH>NH; and loss anticipation versus loss gain outcome, **A5**:BL>N versus **F7**:BLH>NH. We provide these for reference online https://neurovault.org/collections/JVXLTPHC/. Notably, in a direct comparison of the A2: BW>N versus A5: BL>N signal we find no differences in VS or Insula as a function of valence.

With respect to the direct observation of the BOLD signal, we find appropriate separation in anticipation of Big Win and Neutral cues (Figure 4) and signal-to-noise ratio in the VS region

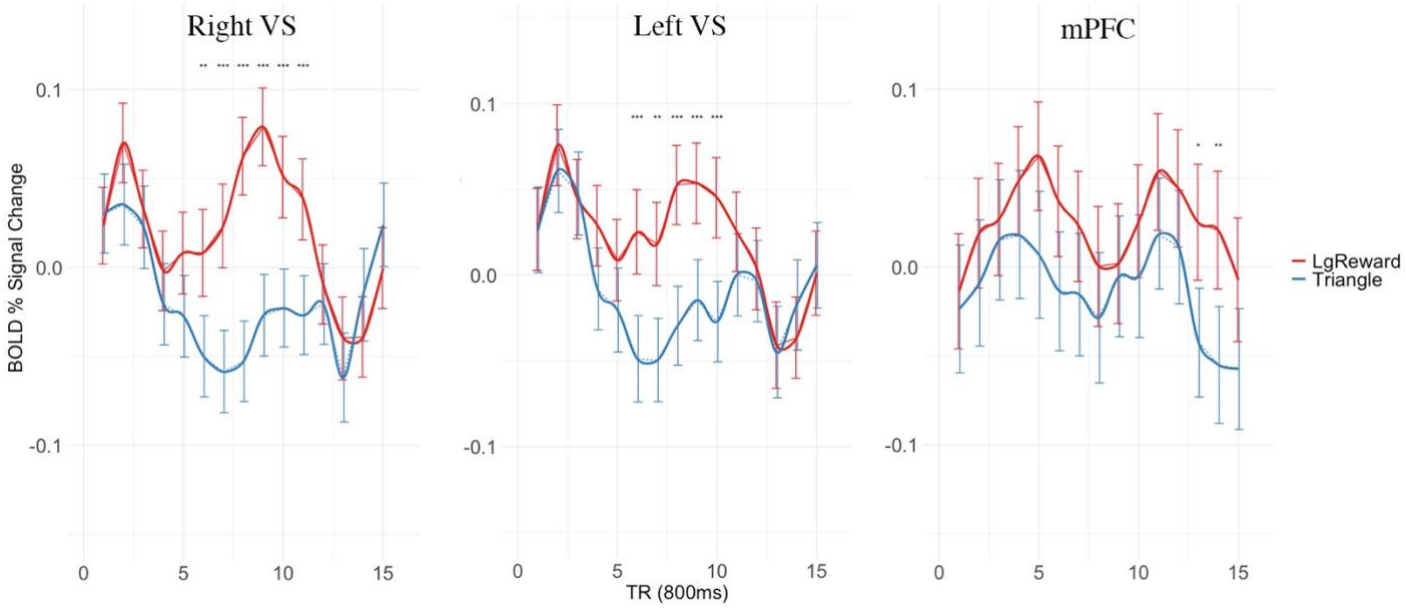

Figure 4: Direct Observation of BOLD signal locked to Cue Onset for Big Win (LgReward) and Neutral (Triangle) for 15 TRs (12 seconds) after Cue Onset
mPFC = medial prefrontal cortex; VS = Ventral Striatum.
Error bars = bootstrapped 90% Confidence Interval; $p < .05$*; $p < .01$**; $p < .001$**

(Supplementary Figure S3). With respect to the anticipation phase, we see the expected peak in BOLD separation between Big Win and Neutral cues around 7-8seconds after cue onset (Figure 4). Such that, this separation is significant from TR 6 ($p < .01$) to TR 11 ($p < .001$) in the Right VS, and TR 6 ($p < .001$) to TR 10 ($p < .001$) in the Left VS, before the undershoot at TR 14. This separation, as expected, does not occur in the mPFC. The nature of the anticipation signal bleeding into the feedback phase is apparent in the bilateral VS when the anticipation cues are locked to the feedback phase (Supplementary Figure S8). Specifically, there is significant separation for the first 4-5 TRs (or 3-4 sec) in the feedback phase in the Big Win as compared to the Neutral phase, until they reverse by TR 10. Since the signal is not appropriately deconvolved in the feedback phase, one approach is to model based on combinations of Hit/Miss trials. In our main feedback contrasts, F6 and F7, we modeled the Big Win versus Neutral Hit, which still

593 demonstrates poor deconvolution in the VS regions (see Supplementary Figure S6). One
594 alternative approach, which we did not model in the whole brain contrasts, is the contrast of Big
595 Win/Loss Hit versus Big Win/Loss Miss. However, direct observation of the BOLD signal
596 (supplementary Figure S7) demonstrates that for Big Win Hit and Big Win Miss, these are nearly
597 identical in the VS BOLD signal. However, whereas the mPFC demonstrates peak separation at
598 TR 14 (~11 sec), this is occurring well into the subsequent trial, it is unclear what this change
599 represents. Overall, we find appropriate peak in direct BOLD signal after anticipation cue onset,
600 but a complicated picture forms in the outcome phase with respect to bilateral VS and mPFC.
601

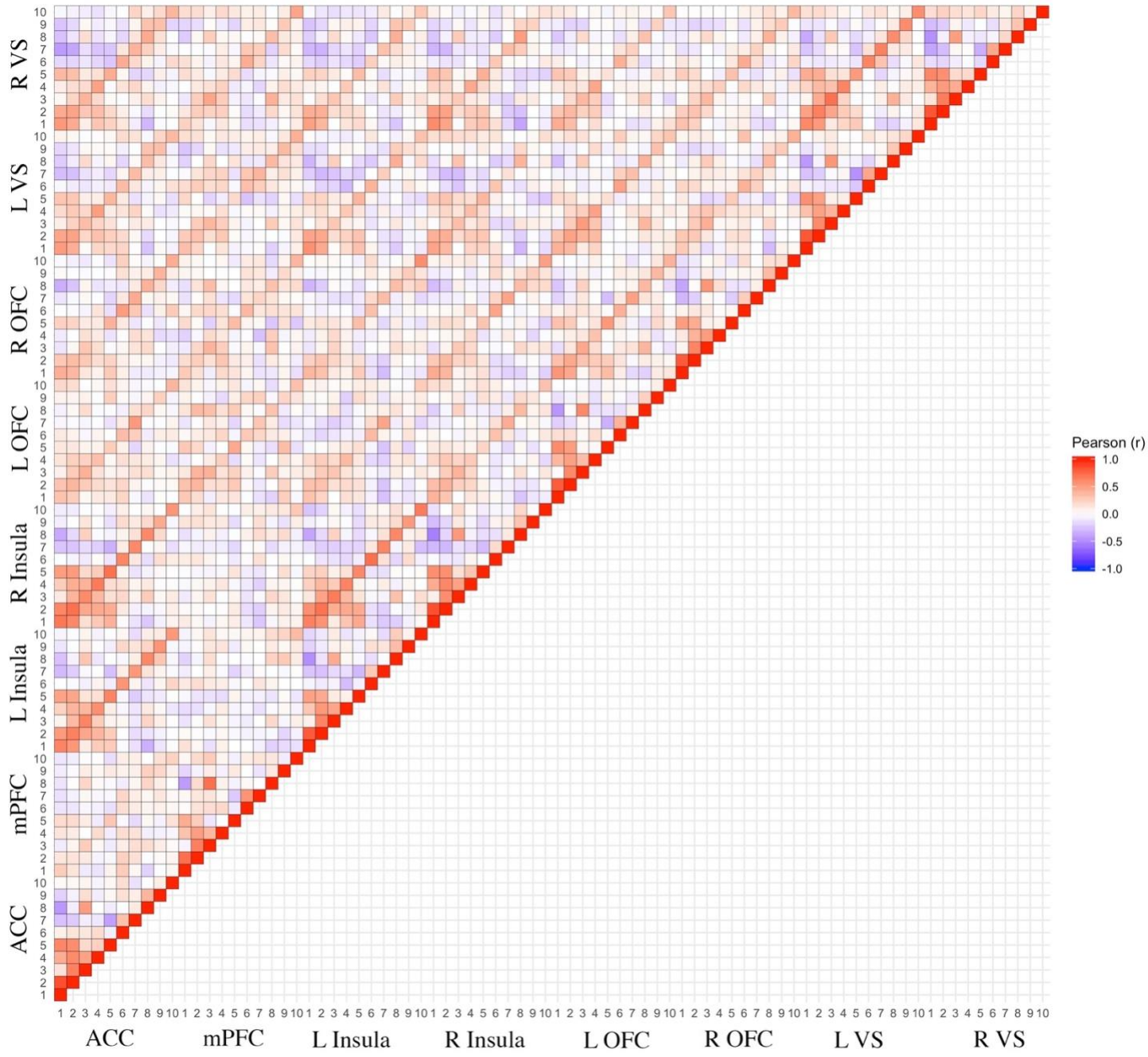

Figure 2: Pearson correlation matrix of 10 contrasts by 8 ROI's Color bar represents associated Pearson *r* value between the 10mm ROI of across 10 contrasts. See Table 1 for associated contrast information. R = Right; L = Left; VS = Ventral Striatum; OFC = Orbitofrontal Cortex; mPFC = medial Prefrontal Cortex; ACC = Anterior Cingulate Cortex

602

603

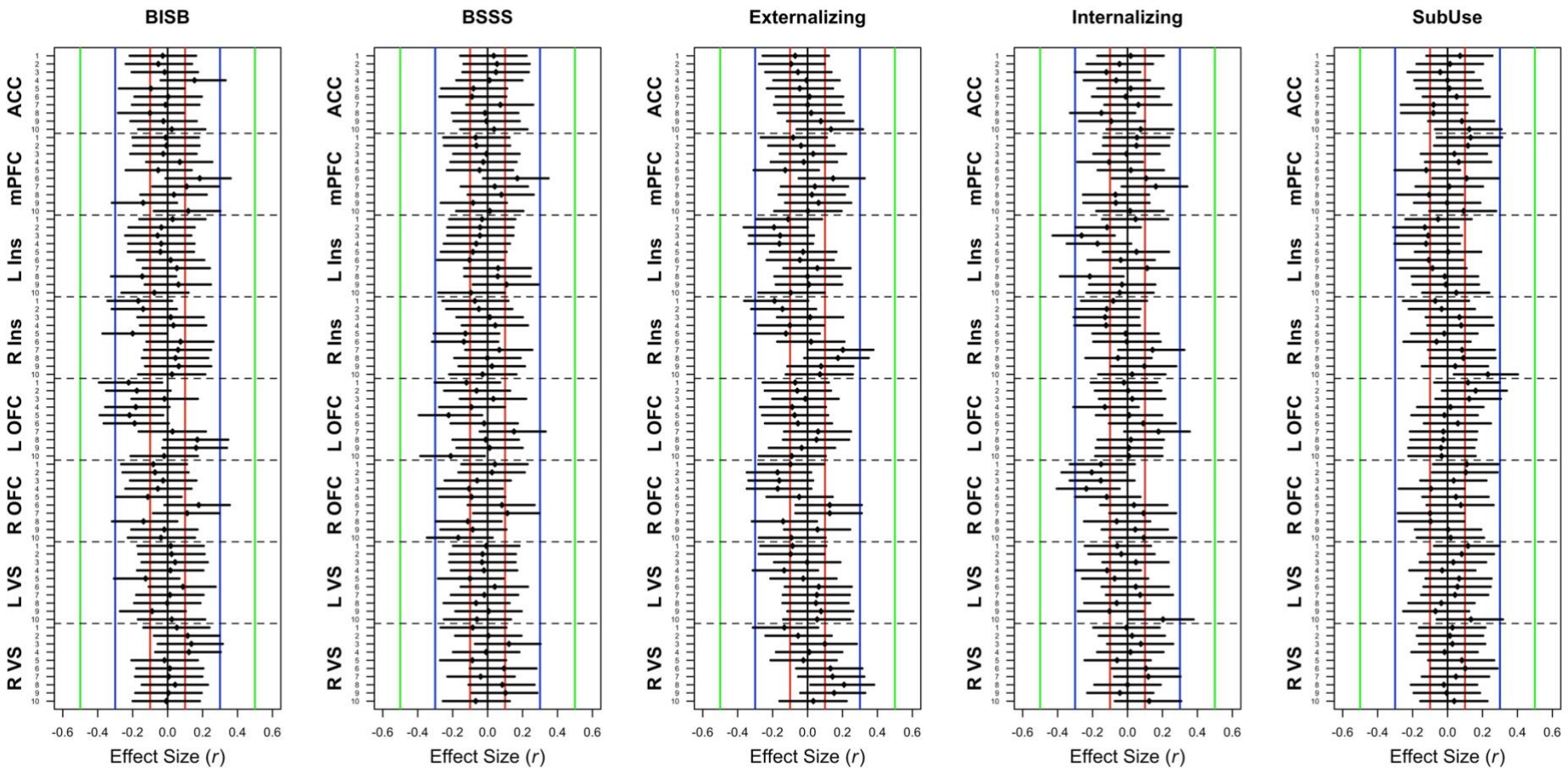

Figure 3. Forest plots displaying the most likely Pearson's *r* value (black diamonds) and 95% Bayesian credible interval (black lines) for correlational relationships between ROI activation estimates from each contrast and behavioral criterion measures. Red, blue and green lines denote "small" (*r*=.10), "moderate" (*r*=.30) and "large" (*r*=.50) effect sizes.

1-10 = Ten contrasts listed in Table 1; ACC = anterior cingulate cortex; mPFC = medial prefrontal cortex; Ins = insula; OFC = orbitofrontal cortex; VS = ventral striatum; L = left; R = right; SubUse = substance use composite measure; BISB = Barratt Impulsiveness Scale-Brief; BSSS = Brief Sensation Seeking Scale

5 Discussion

In this study of the MID task, we performed a comprehensive evaluation of between-contrast similarities and differences in activation maps, mean ROI signal intensity and correlations between ROI activations and behavioral criterion measures. After identifying ten candidate contrasts that have precedent in the previous literature, our within study comparison provides the first look at the comparative differences between these common MID task contrasts. The findings demonstrate similarity between positively and negatively arousing anticipation cues, distinct deactivation of striatal regions during the outcome phase, dissimilarity between Big Win > Small Win anticipation and other anticipation effects, and a relatively weak associations between MID task activations and real-world behaviors. These findings are consistent with previously reported MID task-specific conceptual findings (Bjork et al., 2010) and also have implications for task-general theoretical problems (Hedge et al., 2018; Poldrack & Yarkoni, 2016)

A relatively similar pattern of activation was observed during the Big Win anticipation and the Big Loss anticipation phase. A direct comparison of Big Win and Big Loss anticipation phases revealed negligible differences between the activation in the NAcc and insula in the group level activation maps, and only a small Win-related increase in activation in the mPFC. This similarity in activation profiles during anticipation of both positive and negative stimuli is consistent with a recent meta-analysis demonstrating that approach and avoidance behavior have considerable overlap in activation (Oldham et al., 2018), and prior studies reporting similar activation patterns in young adults (Joseph et al., 2015; Joseph et al., 2016) and substance use populations (Bjork et al., 2008). Although prior models suggested that levels of uncertainty, depending on positive or negative arousal, would elicit activation in the NAcc or insula (Knutson

& Greer, 2008), these difference were not apparent during the anticipation phase. In the direct comparison of BOLD in cue onset of Big Win and Big Loss, although there are subtle differences in our larger sample, these differences are relatively small. This may suggest alternative cognitive processes (such as attention or motivation) that may be involved during the anticipation phase (Abler et al., 2006; Breckel et al., 2011; Krebs et al., 2012; Schouppe et al., 2014), as the NAcc may facilitate detection and attention to cues (Peters et al., 2011), as it serves as a limbic-motor interface that converts signals into action (Floresco, 2015). Thus, the proactive nature of the anticipation phase in the task may permit the individual to strategically prepare via the confounded relationship between attention and reward (Pessoa, 2015).

Our within-task analysis revealed some dissimilarity within contrasts in the anticipation and distinct patterns of activation in the outcome phase compared to the anticipation phase. Although the Big Win versus Small Win contrast activated striatal regions, the contrast was distinct from all others, demonstrating a limited association with other contrasts in the anticipation phase and not apparent improvement in estimates of behavior. Since the contrast is used in the literature (Büchel et al., 2017), more theoretical support is needed to contextualize this modeling approach. Conversely, our comparison of positively and negatively valenced reward outcomes revealed widespread *deactivation* throughout the brain during the outcome phase. These patterns were counter to a recent meta-analysis, using activation likelihood estimation (based on nine studies), that reported no activation in the contrasts of Big Win hit versus Neutral hit conditions in the MID task (Oldham et al., 2018). However, the pattern of deactivation may have been overlooked in a meta-analytic strategy, especially since the technique focuses on reported points of positive activation. Although the pattern of deactivation is consistent in both positively and negatively valenced outcomes, there appears to be a greater

deactivation in the negatively valenced cues as compared to positively valenced cues. This pattern of deactivation is consistent when comparing contrasts modeling the anticipation and outcome phase, whereby there is a consistent negative association between correlations of ROI activation estimates. This suggests a unique pattern of deactivation in the outcome phase that is anti-correlated with the anticipation phase, which is important to consider as the direction of activation may differ as a function of the anticipation and outcome phase, and so should be considered when interpreting an increase or decrease in the MID task. However, since the outcome phase is not temporally-separated from the anticipation phase in most versions of the MID task (Bjork et al., 2010), it is still unclear whether the negative activation in this version is related to some overlap in the BOLD response occurring during the anticipation, probe and outcome phase.

Alternative contrasts may have been examined in the outcome phase, such as Big Win Hit versus Big Win Miss or more complicated contrasts (Bjork et al., 2011; Veroude et al., 2016), but these contrasts would likely still suffer from a signal that is not temporally separated between the outcome and anticipation phase or offer an inadequately powered number of trials. In the direct observation of Big Win Hit versus Big Win Miss cues, we observe that they are nearly identical in during the feedback phase in the VS regions. Although the mPFC demonstrates some separation in these cues, it occurs well into the subsequent trials, which would make it difficult to conclude on the effect. In a version of the MID task that used a filler between the probe and outcome phase, some increases in activation were reported to winning rewards but not loss avoidance (Bjork et al., 2010). Since deactivation was not reported, it is still unclear whether deactivation found here is confounded by contrast type in the outcome phase and/or overlap in BOLD signal. Although it is difficult to do for the current MID design, future

673 designs would benefit from deconvolving the BOLD signal between the anticipation and
674 outcome phase to disentangle the complexity of the design and subsequent conclusions.
675 In addition to understanding the variability across different contrasts of the MID task, it is
676 also critical to consider how patterns of activation across task phases/conditions relate to
677 behaviors, since the task is used in a broad clinical and behavioral literature. In our analysis
678 using psychosocial and clinical criterion measures, we found limited evidence for associations
679 with activations across different phases and conditions. Specifically, the majority of associations
680 between neural activation during the MID task and behavior were likely to be relatively small or
681 negligible effects. As the original design focused on clinical populations (Brian Knutson &
682 Heinz, 2015) and reviews have suggest a robust role of limbic regions in substance use (Balodis
683 & Potenza, 2015) and psychosis (Radua et al., 2015), this may contribute to the weak effects
684 found in our young adult community sample. Although we cannot rule out that this lack of
685 robust associations with behavior may have been due to features of our sample or measures, it
686 stands in stark contrast to the large array of previous studies reporting associations of MID task
687 activations with various real-world outcomes (Büchel et al., 2017; Boecker et al., 2014).
688 However, our findings are broadly consistent with recent work that has reported a distinct
689 contrast between the effects found in studies with (median r = .16) and without preregistration
690 (median r = .36) respectively) (Schäfer & Schwarz, 2019) and with findings in large, diverse data
691 sets which indicate that neuroimaging markers often explain only very small portions of the
692 variance in behavioral outcomes of interest (Nees et al., 2012; Paulus et al., 2019; Paulus &
693 Thompson, 2019). This has led some (Paulus and Thompson, 2019) to suggest that small effects
694 are the “new normal” in clinical neuroscience research.

One reason for discrepancy between our results and prior reports of more robust MID task associations with behavior is that effect sizes may have been overestimated in previous studies with smaller samples. Until recently, neuroimaging studies of individual differences have frequently been underpowered (Cremers et al., 2017; Yarkoni, 2009), which tends to cause the size and replicability of effects to be dramatically overestimated due to a combination of noise in small samples and the "statistical significance filter" (Gelman & Loken, 2014; Vasishth et al., 2018). Hence, when viewed in the broader context of current challenges faced by neuroimaging studies of individual differences, our findings suggest that researchers should be prepared for relationships between MID task activations and clinical or real-world outcomes of interest to be of small size, and design their studies accordingly. For example, the use of large data sets from collaborative efforts (e.g., ABCD: Casey et al., 2018) may be preferable to smaller samples collected by individual labs (Beltz & Weigard, 2019; Paulus & Thompson, 2019).

Beyond the possibility that effect sizes in previous MID studies may have simply been inflated, the lack of relationships may also be attributed to problematic validity of fMRI-based tasks and the underlying assumptions in the cognitive processes and their related phenomena, such as positive or negative valence. A large proportion of tasks in fMRI are experiment based, whereby conditions are manipulated to evoke excitation of a specific cognitive processes. Although the MID task evokes distinct neural processes that are consistent with current conceptualizations of the mesolimbic system (Knutson & Greer, 2008) – which we find distinct separation in VS to valence – the classic metric of validity, that a test measures what it claims to measure (Cronbach & Meehl, 1955; Kelley, 1927), is underexplored. In fMRI studies of individual variation, such as behavioral differences that may be associated with neural measures of reward, requires the combination of experimental and correlation work which has been

718 considered to be two distinct traditions in psychology (Cronbach, 1957). Correlation work
719 attempts to increase between individual variation, whereas experimental work attempts to limit
720 the between-individual variation; the latter methodological practice has been argued to contribute
721 to the poor predictive effect of cognitive measures in correlational research (Dang et al., 2020;
722 Hedge et al., 2018). Together, the weak predictive effect of cognitive tasks and poor test re-test
723 of fMRI (Elliott et al., 2020) can contribute to the unreliable estimates of different task contrasts.
724 Moreover, the inferential processes in task-based fMRI pose conceptual challenges. It has
725 been argued that the standard approaches in task-based fMRI, that utilize the technique of
726 subtracting conditions, are fundamentally flawed in achieving the isolated mental functions in
727 neural substrates, making it difficult to map brain to behavior (Poldrack & Yarkoni, 2016).
728 Poldrack & Yarkoni (2016) recommended that there are basic conceptual difficulties in task-
729 based fMRI 'that remain widely underappreciated within the neuroimaging community' (pg.
730 589). This is observed in the MID task, as *conceptually* the subtraction intends to measure
731 approach and avoidance of positive and negative conditions (Knutson & Greer, 2008), but, this is
732 not consistent in the activation patterns of valence (insula) and approach (NAcc) structures that,
733 here, are activated similarly in both. Although using monetary value illustrates control of
734 magnitude, probability and timing (Knutson & Greer, 2008), adding a discrete step with positive
735 or negative monetary cues may not be sophisticated enough to identify valence and approach
736 over and above processes of attention and/or motivation within an individual. This poor one-to-
737 one mapping makes it especially difficult to predict socioemotional vulnerabilities due to the
738 heterogeneity of populations that do not uniquely fit into one-to-one neural mappings (Paulus &
739 Thompson, 2019). While the MID task poses to measure distinct positive and negative valenced
740 systems in two distinct phases (Balodis & Potenza, 2015; Knutson et al., 2000; Knutson & Greer,

2008; Oldham et al., 2018), the nature to which these phenomena vary or are consistent have not been validated. And in fact, our work in a community sample of young adults suggests that they may not significantly differ in terms of the structures that are involved.

Therefore, as has long been acknowledged in behavioral research (Cronbach & Meehl, 1955), it is crucial to test whether activations from fMRI contrasts cluster in the same manner in which they are proposed to by theorists and task designers. Likewise, if reviewers are to contrast between studies and tasks at the results level, implying a construct of a cognitive process, it would be valuable to the field to know whether within- and between-task conditions converge on a hypothesized phenomenon (Campbell & Fiske, 1959). Until assumptions are considered in cognitive processes of fMRI tasks, simple manipulations that are laden with theoretical claims of reward processes will provide overconfidence in the robustness of findings (Poldrack & Yarkoni, 2016; Vasishth et al., 2018). Due to the variability in our findings, the current state of practices used in the MID task (and potentially others) will inevitably result in playing '20 questions with nature' (Newell, 1973), which will result in some winners and some losers (Button et al., 2013). By modeling multiple versions of an unvalidated phenomena in contrasts, the incorrect measure of error will invariability result in significance 5-percent of the time (Fisher, 1926), and in small samples, this may be lauded as a significant finding between neural substrates of reward and behavior, which can alter designs and conclusions of subsequent studies (Button et al., 2013).

Although our findings suggest a high level of variability between contrast choices and behavioral associations, several measures can be taken to improve the reproducibility of results in the MID task literature. A first and immediate step that can been taken by researchers is increasing sample sizes of samples using fMRI. Currently, a large proportion of fMRI studies are substantially underpowered for finding the effect they are testing (Szucs & Ioannidis, 2017). In

addition to improving power, researchers would benefit from assessing how the MID contrast values fit in a larger nomological network of neural and behavioral constructs, beyond an abstract subtraction processes that presumes a process of motivation or consumption of reward (Poldrack & Yarkoni, 2016). One approach may be to use parametric modulators, which has been used in prior analyses, but is largely underutilized (Aloi et al., 2019; Joseph et al., 2016). In addition to improving estimates of functional parcels (Nikolaidis et al., 2020), multivariate pattern analyses may help with the reproducibility of theorized cognitive processes (Hong et al., 2019). Multivariate, cross-validated, pattern analyses can provide *a priori* activation patterns and locations that can be confirmed out of sample, reducing the possibility of exploring multiple hypotheses. Furthermore, in order to characterize individual variability in neural models, researchers should implement functional organization techniques to explain changes in behavior and cognitive processes (Beltz et al., 2016; Greene et al., 2018; Yip et al., 2019; Zhang et al., 2019). For example, Zhang and colleagues (2019) used a network modeling approach to identify a developmentally stable architecture of emotion related findings, providing some reliable estimates. Further, the network models of task-based fMRI may aid researchers in uncovering the neural architecture of cognitive processes (Medaglia et al., 2015), such that connectivity metrics may provide predictive effects of individual traits (Greene et al., 2018). By using individual and group level estimates of connectivity patterns (Beltz et al., 2016), task-based analyses may improve the identification and replication of neural signatures that will aid researchers studying developmental and clinical differences (Yip et al., 2019; Zhang et al., 2019).

5.1 Limitations

Although the findings here pose significant implications, there are multiple limitations. First, the nature of our findings are specific only a modified version of MID task that was administered in a young adult sample, so the implications should be considered and confirmed in a separate sample(s) to determine which effects are converge between samples and which are limited to a sample. Second, the correlates between ROI activation and self-reported behavior may be underestimated, such that behavior that is collected contemporaneously with the scan acquisition or in the nature that brain predicts behavior may produce different effects. Third, only a subset of common *a priori* contrasts were selected from the literature. Alternative contrasts, such as the linear combination of winning or alternative contrasts during the outcome phase, should be considered in future work. Further, since the anticipation and outcome phase in this task were not jittered, we could not directly contrast these phases at the individual level (only group level), due to risk of collinearity. Finally, due to the outcome phase containing variable number of trials as a function of 60% accuracy rate, the activation patterns may be influenced by the surprise of the event(s) (Vassena et al., 2020), which should be considered in future work.

It is worth noting, that some of differences between positive and negative cues in our and previous studies may depend on age-related factors and sample characteristics. For instance, while our results did not demonstrate a meaningful difference in the activation of the VS or insula between big win and big lose anticipation phases, age related differences have been previously reported using this task (Bjork et al., 2010; Cope et al., 2019), such that increases in activation during big win anticipation trials were greater in older adults (Bjork et al., 2010), and reduced activation in reponse to lose big anticipation in 9-12 year old's (Cope et al., 2019). This suggests patterns of activation during the MID task within and between sample comparisons has be considered when age-related effects are present, as qualitative differences between some

809 contrasts may not be easily apparent. Furthermore, whereas these analyses focus on a
810 community-recruited young adult sample, previous reviews focused on clinical population
811 (Balodis & Potenza, 2015; Radua et al., 2015), and these results should be considered in the
812 future within a clinical population to assess how associations would change in light of clinical
813 factors.

814 Finally, in this analysis we were not able to explore at what degree individual intrinsic
815 motivation differentiated across cue types (win/loss/neutral). Although the accuracy in the task
816 was used as a marker, it is difficult to determine how interested a participant was in the task
817 conditions. Future work should consider how relationships can be accounted for by self-reported
818 metrics, and whether the degree of task-negative activation in the default mode network may be a
819 useful indicator of vigor (or attention) during the MID task.

820 ## 5.2 Conclusion

821 Although the MID task has been used to measure neural substrates of reward processing,
822 modeling techniques have varied substantially between studies. While the structure of the task
823 has been proposed to measure varying levels of arousal and valence, it is still unclear whether
824 findings from different within task comparisons can be easily generalized between studies. Our
825 comparison of within-sample MID task contrasts during multiband fMRI revealed more
826 similarities than differences between positive and negative cues during the anticipation contrast,
827 dissimilarity of a distinct contrasts during the anticipation phase, a robust deactivation effect in
828 the outcome phase, and behavioral associations that are less robust than previously thought.
829 These findings point to the need for caution in future work that make attempts at generalization,
830 and encourage researchers to power their studies for effects that may be smaller than previously
831 hypothesized.

References

Abler, B., Walter, H., Erk, S., Kammerer, H., & Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage*, *31*(2), 790–795. https://doi.org/10.1016/j.neuroimage.2006.01.001

Achenbach, T. M. (2013). *DSM-Oriented Guide for the Achenbach System of Empirically Based Assessment (ASEBA).* Burlington, VT: University of Vermont Research Center for Children, Youth and Families. https://trove.nla.gov.au/version/11167240

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment*. Burlington, VT : ASEBA. https://trove.nla.gov.au/version/11167240

Aloi, J., Meffert, H., White, S. F., Blair, K. S., Hwang, S., Tyler, P. M., Thornton, L. C., Crum, K. I., Adams, K. O., Killanin, A. D., Filbey, F., Pope, K., & Blair, R. J. R. (2019). Differential dysfunctions related to alcohol and cannabis use disorder symptoms in reward and error-processing neuro-circuitries in adolescents. *Developmental Cognitive Neuroscience*, *36*, 100618. https://doi.org/10.1016/j.dcn.2019.100618

Balleine, B. W., & O'Doherty, J. P. (2010). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology*, *35*(1), 48–69. https://doi.org/10.1038/npp.2009.131

Balodis, I. M., & Potenza, M. N. (2015). Anticipatory Reward Processing in Addicted Populations: A Focus on the Monetary Incentive Delay Task. *Biological Psychiatry*, *77*(5), 434–444. https://doi.org/10.1016/j.biopsych.2014.08.020

Beltz, A. M., & Weigard, A. (2019). Methodological Advances in Leveraging Neuroimaging Datasets in Adolescent Substance Use Research. *Current Addiction Reports*, *6*(4), 495–503. https://doi.org/10.1007/s40429-019-00275-x

Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the Nomothetic and Idiographic Approaches to the Analysis of Clinical Data. *Assessment*, *23*(4), 447–458. https://doi.org/10.1177/1073191116648209

Berridge, K. C. (2007). The debate over dopamine's role in reward: The case for incentive salience. *Psychopharmacology*, *191*(3), 391–431. https://doi.org/10.1007/s00213-006-0578-x

Berridge, K. C. (2019). Affective valence in the brain: Modules or modes? *Nature Reviews Neuroscience*, *20*(4), 225–234. https://doi.org/10.1038/s41583-019-0122-8

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure Systems in the Brain. *Neuron*, *86*(3), 646–664. https://doi.org/10.1016/j.neuron.2015.02.018

Berridge, K. C., & O'Doherty, J. P. (2014). Chapter 18—From Experienced Utility to Decision Utility. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 335–351). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00018-8

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences*, *26*(9), 507–513. https://doi.org/10.1016/S0166-2236(03)00233-9

Bjork, J. M., Smith, A. R., Chen, G., & Hommer, D. W. (2010). Adolescents, Adults and Rewards: Comparing Motivational Neurocircuitry Recruitment Using fMRI. *PLOS ONE*, *5*(7), e11440. https://doi.org/10.1371/journal.pone.0011440

Bjork, J. M., Smith, A. R., Chen, G., & Hommer, D. W. (2011). Mesolimbic recruitment by nondrug rewards in detoxified alcoholics: Effort anticipation, reward anticipation, and

reward delivery. *Human Brain Mapping*, *33*(9), 2174–2188. https://doi.org/10.1002/hbm.21351

Bjork, J. M., Smith, A. R., & Hommer, D. W. (2008). Striatal sensitivity to reward deliveries and omissions in substance dependent patients. *NeuroImage*, *42*(4), 1609–1621. https://doi.org/10.1016/j.neuroimage.2008.06.035

Boecker, R., Holz, N. E., Buchmann, A. F., Blomeyer, D., Plichta, M. M., Wolf, I., Baumeister, S., Meyer-Lindenberg, A., Banaschewski, T., Brandeis, D., & Laucht, M. (2014). Impact of early life adversity on reward processing in young adults: EEG-fMRI results from a prospective study over 25 years. *PloS One*, *9*(8), e104185. https://doi.org/10.1371/journal.pone.0104185

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 1–7. https://doi.org/10.1038/s41586-020-2314-9

Bourque, J., Spechler, P. A., Potvin, S., Whelan, R., Banaschewski, T., Bokde, A. L. W., Bromberg, U., Büchel, C., Quinlan, E. B., Desrivières, S., Flor, H., Frouin, V., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Paillère-Martinot, M.-L., McEwen, S. C., Nees, F., … IMAGEN Consortium. (2017). Functional Neuroimaging Predictors of Self-Reported Psychotic Symptoms in Adolescents. *The American Journal of Psychiatry*, *174*(6), 566–575. https://doi.org/10.1176/appi.ajp.2017.16080897

Breckel, T. P. K., Giessing, C., & Thiel, C. M. (2011). Impact of brain networks involved in vigilance on processing irrelevant visual motion. *NeuroImage*, *55*(4), 1754–1762. https://doi.org/10.1016/j.neuroimage.2011.01.025

Breiter, H. C., Berke, J. D., Kennedy, W. A., Rosen, B. R., & Hyman, S. E. (1996). Activation of striatum and amygdala during reward conditioning: An FMRI study. *NeuroImage*, *3 Supplement*(3), S220. https://doi.org/10.1016/S1053-8119(96)80222-9

Büchel, C., Peters, J., Banaschewski, T., Bokde, A. L. W., Bromberg, U., Conrod, P. J., Flor, H., Papadopoulos, D., Garavan, H., Gowland, P., Heinz, A., Walter, H., Ittermann, B., Mann, K., Martinot, J.-L., Paillère-Martinot, M.-L., Nees, F., Paus, T., Pausova, Z., … Knutson, B. (2017). Blunted ventral striatal responses to anticipated rewards foreshadow problematic drug use in novelty-seeking adolescents. *Nature Communications*, *8*(1), 1–11. https://doi.org/10.1038/ncomms14140

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. https://doi.org/10.1037/h0046016

Cao, Z., Bennett, M., O'Halloran, L., Pragulbickaite, G., Flanagan, L., McHugh, L., & Whelan, R. (n.d.). Aberrant reward prediction errors in young adult at-risk alcohol users. *Addiction Biology*, *n/a*(n/a), e12873. https://doi.org/10.1111/adb.12873

Cao, Z., Bennett, M., Orr, C., Icke, I., Banaschewski, T., Barker, G. J., Bokde, A. L. W., Bromberg, U., Büchel, C., Quinlan, E. B., Desrivières, S., Flor, H., Frouin, V., Garavan, H., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Nees, F., … Whelan, R. (2019). Mapping adolescent reward anticipation, receipt, and prediction error during the monetary incentive delay task. *Human Brain Mapping*, *40*(1), 262–283. https://doi.org/10.1002/hbm.24370

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., … ABCD Imaging Acquisition Workgroup. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Chan, R. C. K., Li, Z., Li, K., Zeng, Y.-W., Xie, W.-Z., Yan, C., Cheung, E. F. C., & Jin, Z. (2016). Distinct processing of social and monetary rewards in late adolescents with trait anhedonia. *Neuropsychology*, *30*(3), 274–280. https://doi.org/10.1037/neu0000233

Colich, N. L., Ho, T. C., Ellwood-Lowe, M. E., Foland-Ross, L. C., Sacchet, M. D., LeMoult, J. L., & Gotlib, I. H. (2017). Like mother like daughter: Putamen activation as a mechanism underlying intergenerational risk for depression. *Social Cognitive and Affective Neuroscience*, *12*(9), 1480–1489. https://doi.org/10.1093/scan/nsx073

Cope, L. M., Martz, M. E., Hardee, J. E., Zucker, R. A., & Heitzeg, M. M. (2019). Reward activation in childhood predicts adolescent substance use initiation in a high-risk sample. *Drug and Alcohol Dependence*, *194*, 318–325. https://doi.org/10.1016/j.drugalcdep.2018.11.003

942 Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and
943 inference in fMRI. *PLOS ONE*, *12*(11), e0184923.
944 https://doi.org/10.1371/journal.pone.0184923
945 Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*,
946 *12*(11), 671–684. https://doi.org/10.1037/h0043943
947 Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological*
948 *Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957
949 Dang, J., King, K. M., & Inzlicht, M. (2020). Why Are Self-Report and Behavioral Measures
950 Weakly Correlated? *Trends in Cognitive Sciences*.
951 https://doi.org/10.1016/j.tics.2020.01.007
952 Demidenko, M. I., Huntley, E., Jahn, A. A., Thomason, M., Monk, C. S., & Keating, D. (2019).
953 *Cortical and subcortical response to the anticipation of reward in high and average risk-*
954 *taking adolescents*. https://doi.org/10.31234/osf.io/fnpvx
955 Dhingra, I., Zhang, S., Zhornitsky, S., Le, T. M., Wang, W., Chao, H. H., Levy, I., & Li, C.-S. R.
956 (2019). The effects of age on reward magnitude processing in the monetary incentive
957 delay task. *NeuroImage*, 116368. https://doi.org/10.1016/j.neuroimage.2019.116368
958 Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*,
959 *26*(3), 297–302. JSTOR. https://doi.org/10.2307/1932409
960 DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science*, *11*(3),
961 189–228. https://doi.org/10.1214/ss/1032280214
962 Dugré, J. R., Dumais, A., Bitar, N., & Potvin, S. (2018). Loss anticipation and outcome during
963 the Monetary Incentive Delay Task: A neuroimaging systematic review and meta-
964 analysis. *PeerJ*, *6*, e4749. https://doi.org/10.7717/peerj.4749

965 Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L.,
966 Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What Is the Test-Retest Reliability of
967 Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-
968 Analysis: *Psychological Science*. https://doi.org/10.1177/0956797620916786
969 Ernst, M., & Luciana, M. (2015). Neuroimaging of the dopamine/reward system in adolescent
970 drug use. *CNS Spectrums*, *20*(4), 427–441. https://doi.org/10.1017/S1092852915000395
971 Fisher, R. A. (1926). *048: The Arrangement of Field Experiments.*
972 https://digital.library.adelaide.edu.au/dspace/handle/2440/15191
973 Floresco, S. B. (2015). The nucleus accumbens: An interface between cognition, emotion, and
974 action. *Annual Review of Psychology*, *66*, 25–52. https://doi.org/10.1146/annurev-psych-
975 010213-115159
976 Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis—A"
977 garden of forking paths"—Explains why many statistically significant comparisons don't
978 hold up. *American Scientist*, *102*(6), 460–466.
979 Gonzalez, M. Z., Allen, J. P., & Coan, J. A. (2016). Lower neighborhood quality in adolescence
980 predicts higher mesolimbic sensitivity to reward anticipation in adulthood.
981 *Developmental Cognitive Neuroscience*, *22*, 48–57.
982 https://doi.org/10.1016/j.dcn.2016.10.003
983 Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state
984 manipulation improves prediction of individual traits. *Nature Communications*, *9*(1), 1–
985 13. https://doi.org/10.1038/s41467-018-04920-3
986 Haber, S. N., & Knutson, B. (2010). The Reward Circuit: Linking Primate Anatomy and Human
987 Imaging. *Neuropsychopharmacology*, *35*(1), 4–26. https://doi.org/10.1038/npp.2009.129

Hagler, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., Sutherland, M. T., Casey, B. J., Barch, D. M., Harms, M. P., Watts, R., Bjork, J. M., Garavan, H. P., Hilmer, L., Pung, C. J., Sicat, C. S., Kuperman, J., Bartsch, H., Xue, F., … Dale, A. M. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage*, *202*, 116091. https://doi.org/10.1016/j.neuroimage.2019.116091

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hong, Y.-W., Yoo, Y., Han, J., Wager, T. D., & Woo, C.-W. (2019). False-positive neuroimaging: Undisclosed flexibility in testing spatial hypotheses allows presenting anything as a replicated finding. *NeuroImage*, *195*, 384–395. https://doi.org/10.1016/j.neuroimage.2019.03.070

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences*, *32*(3), 401–414. https://doi.org/10.1016/S0191-8869(01)00032-0

JASP Team. (2019). *ASP (version 0.10. 2)[computer software]*. https://jasp-stats.org

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156.

Johnston, L. D., Miech, R. A., O'Malley, P. M., Bachman, J. g, Schulenberg, J. E., & Patrick, M. E. (2018). *Monitoring the Future National Survey Results on Drug Use, 1975-2017: Overview, Key Findings on Adolescent Drug Use*. Institute for Social Research. https://eric.ed.gov/?id=ED589762

Joseph, J. E., Zhu, X., Lynam, D., & Kelly, T. H. (2016). Modulation of meso-limbic reward processing by motivational tendencies in young adolescents and adults. *NeuroImage*, *129*, 40–54. https://doi.org/10.1016/j.neuroimage.2015.12.005

Joseph, J., Zhu, X., Benca, C., Baik, G., Davies, F., & Kelly, T. H. (2015). Adolescents are driven by incentive valence, not magnitude, on the monetary incentive delay task. *Drug and Alcohol Dependence*, *146*, e151. https://doi.org/10.1016/j.drugalcdep.2014.09.328

Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Co.

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: Characterization with rapid event-related fMRI. *NeuroImage*, *18*(2), 263–272. https://doi.org/10.1016/s1053-8119(02)00057-5

Knutson, B., & Greer, S. (2008). Anticipatory affect: Neural correlates and consequences for choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1511), 3771–3786. https://doi.org/10.1098/rstb.2008.0155

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage*, *12*(1), 20–27. https://doi.org/10.1006/nimg.2000.0593

Knutson, Brian, & Heinz, A. (2015). Probing Psychiatric Symptoms with the Monetary Incentive Delay Task. *Biological Psychiatry*, *77*(5), 418–420. https://doi.org/10.1016/j.biopsych.2014.12.022

Krebs, R. M., Boehler, C. N., Roberts, K. C., Song, A. W., & Woldorff, M. G. (2012). The Involvement of the Dopaminergic Midbrain and Cortico-Striatal-Thalamic Circuits in the Integration of Reward Prospect and Attentional Task Demands. *Cerebral Cortex*, *22*(3), 607–615. https://doi.org/10.1093/cercor/bhr134

Lamm, C., Benson, B. E., Guyer, A. E., Perez-Edgar, K., Fox, N. A., Pine, D. S., & Ernst, M. (2014). Longitudinal study of striatal activation to reward and loss anticipation from mid-adolescence into late adolescence/early adulthood. *Brain and Cognition*, *89*, 51–60. https://doi.org/10.1016/j.bandc.2013.12.003

Lancaster, T. M., Linden, D. E., Tansey, K. E., Banaschewski, T., Bokde, A. L. W., Bromberg, U., Büchel, C., Cattrell, A., Conrod, P. J., Flor, H., Frouin, V., Gallinat, J., Garavan, H., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Paillère Martinot, M.-L., Artiges, E., … IMAGEN Consortium. (2016). Polygenic Risk of Psychosis and Ventral Striatal Activation During Reward Processing in Healthy Adolescents. *JAMA Psychiatry*, *73*(8), 852–861. https://doi.org/10.1001/jamapsychiatry.2016.1135

Landes, I., Bakos, S., Kohls, G., Bartling, J., Schulte-Körne, G., & Greimel, E. (2018). Altered neural processing of reward and punishment in adolescents with Major Depressive Disorder. *Journal of Affective Disorders*, *232*, 23–33. https://doi.org/10.1016/j.jad.2018.01.017

Lutz, K., & Widmer, M. (2014). What can the monetary incentive delay task tell us about the neural processing of reward and punishment? *Neuroscience and Neuroeconomics*, *3*, 33–45. https://doi.org/10.2147/NAN.S38864

Ly, A., Marsman, M., & Wagenmakers, E.-J. (2018). Analytic posteriors for Pearson's correlation coefficient. *Statistica Neerlandica*, *72*(1), 4–13. https://doi.org/10.1111/stan.12111

Maitra, R. (2010). A re-defined and generalized percent-overlap-of-activation measure for studies of fMRI reproducibility and its use in identifying outlier activation maps. *NeuroImage*, *50*(1), 124–135. https://doi.org/10.1016/j.neuroimage.2009.11.070

Maresh, E. L., Stim, J. J., Van Voorhis, A. C., Kang, S. S., Luciana, M., Sponheim, S. R., & Urošević, S. (2019). Neurophysiological correlates of cognitive control and approach motivation abnormalities in adolescent bipolar disorders. *Cognitive, Affective & Behavioral Neuroscience*, *19*(3), 677–691. https://doi.org/10.3758/s13415-019-00719-x

Martz, M. E., Trucco, E. M., Cope, L. M., Hardee, J. E., Jester, J. M., Zucker, R. A., & Heitzeg, M. M. (2016). Association of Marijuana Use With Blunted Nucleus Accumbens Response to Reward Anticipation. *JAMA Psychiatry*, *73*(8), 838–844. https://doi.org/10.1001/jamapsychiatry.2016.1161

McGonigle, D. J., Howseman, A. M., Athwal, B. S., Friston, K. J., Frackowiak, R. S., & Holmes, A. P. (2000). Variability in fMRI: An examination of intersession differences. *NeuroImage*, *11*(6 Pt 1), 708–734. https://doi.org/10.1006/nimg.2000.0562

Medaglia, J. D., Lynall, M.-E., & Bassett, D. S. (2015). Cognitive network neuroscience. *Journal of Cognitive Neuroscience*, *27*(8), 1471–1491. https://doi.org/10.1162/jocn_a_00810

Mikita, N., Simonoff, E., Pine, D. S., Goodman, R., Artiges, E., Banaschewski, T., Bokde, A. L., Bromberg, U., Büchel, C., Cattrell, A., Conrod, P. J., Desrivières, S., Flor, H., Frouin, V., Gallinat, J., Garavan, H., Heinz, A., Ittermann, B., Jurk, S., … Stringaris, A. (2016). Disentangling the autism-anxiety overlap: FMRI of reward processing in a community-based longitudinal study. *Translational Psychiatry*, *6*(6), e845. https://doi.org/10.1038/tp.2016.107

Mori, A., Okamoto, Y., Okada, G., Takagaki, K., Jinnin, R., Takamura, M., Kobayakawa, M., & Yamawaki, S. (2016). Behavioral activation can normalize neural hypoactivation in subthreshold depression during a monetary incentive delay task. *Journal of Affective Disorders*, *189*, 254–262. https://doi.org/10.1016/j.jad.2015.09.036

Nees, F., Tzschoppe, J., Patrick, C. J., Vollstädt-Klein, S., Steiner, S., Poustka, L., Banaschewski, T., Barker, G. J., Büchel, C., Conrod, P. J., Garavan, H., Heinz, A., Gallinat, J., Lathrop, M., Mann, K., Artiges, E., Paus, T., Poline, J.-B., Robbins, T. W., … IMAGEN Consortium. (2012). Determinants of early alcohol use in healthy adolescents: The differential contribution of neuroimaging and psychological factors. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *37*(4), 986–995. https://doi.org/10.1038/npp.2011.282

Nestor, L. J., Behan, B., Suckling, J., & Garavan, H. (2019). Cannabis-dependent adolescents show differences in global reward-associated network topology: A functional connectomics approach. *Addiction Biology*, e12752. https://doi.org/10.1111/adb.12752

Newell, A. (1973). You can ’t play 20 questions with nature and win: Projective comments on the papers of this symposium. In *Visual Information Processing* (pp. 283–308). Elsevier. https://doi.org/10.1016/B978-0-12-170150-5.50012-3

Nikolaidis, A., Heinsfeld, A. S., Xu, T., Bellec, P., Vogelstein, J., & Milham, M. (2020). Bagging improves reproducibility of functional parcellation of the human brain. *NeuroImage*, 116678. https://doi.org/10.1016/j.neuroimage.2020.116678

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., & Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: A neuroimaging meta-analysis of the monetary incentive delay task. *Human Brain Mapping*, *39*(8), 3398–3418. https://doi.org/10.1002/hbm.24184

Papanastasiou, E., Mouchlianitis, E., Joyce, D. W., McGuire, P., Banaschewski, T., Bokde, A. L. W., Bromberg, U., Büchel, C., Quinlan, E. B., Desrivières, S., Flor, H., Frouin, V., Garavan, H., Spechler, P., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Paillère Martinot, M.-L., … IMAGEN Consortium. (2018). Examination of the Neural Basis of Psychoticlike Experiences in Adolescence During Reward Processing. *JAMA Psychiatry*, *75*(10), 1043–1051. https://doi.org/10.1001/jamapsychiatry.2018.1973

Paulus, M. P., Squeglia, L. M., Bagot, K., Jacobus, J., Kuplicki, R., Breslin, F. J., Bodurka, J., Morris, A. S., Thompson, W. K., Bartsch, H., & Tapert, S. F. (2019). Screen media activity and brain structure in youth: Evidence for diverse structural correlation networks from the ABCD study. *NeuroImage*, *185*, 140–153. https://doi.org/10.1016/j.neuroimage.2018.10.040

Paulus, M. P., & Thompson, W. K. (2019). The Challenges and Opportunities of Small Effects: The New Normal in Academic Psychiatry. *JAMA Psychiatry*, *76*(4), 353–354. https://doi.org/10.1001/jamapsychiatry.2018.4540

Pessoa, L. (2015). Multiple influences of reward on perception and attention. *Visual Cognition*, *23*(1–2), 272–290. https://doi.org/10.1080/13506285.2014.974729

Peters, M. S., Demeter, E., Lustig, C., Bruno, J. P., & Sarter, M. (2011). Enhanced Control of Attention by Stimulating Mesolimbic–Corticopetal Cholinergic Circuitry. *Journal of Neuroscience*, *31*(26), 9760–9771. https://doi.org/10.1523/JNEUROSCI.1902-11.2011

Poldrack, R. A., & Yarkoni, T. (2016). From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure. *Annual Review of Psychology*, *67*(1), 587–612. https://doi.org/10.1146/annurev-psych-122414-033729

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Radua, J., Schmidt, A., Borgwardt, S., Heinz, A., Schlagenhauf, F., McGuire, P., & Fusar-Poli, P. (2015). Ventral Striatal Activation During Reward Processing in Psychosis: A Neurofunctional Meta-Analysis. *JAMA Psychiatry*, *72*(12), 1243–1251. https://doi.org/10.1001/jamapsychiatry.2015.2196

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non reinforcement. In *In AH. Black & W.F. Prokasy (eds.), Classical conditioning II: current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Richards, J. M., Plate, R. C., & Ernst, M. (2013). A systematic review of fMRI reward paradigms used in studies of adolescents vs. adults: The impact of task design and implications for understanding neurodevelopment. *Neuroscience & Biobehavioral Reviews*, *37*(5), 976–991. https://doi.org/10.1016/j.neubiorev.2013.03.004

Risk, B. B., Kociuba, M. C., & Rowe, D. B. (2018). Impacts of simultaneous multislice acquisition on sensitivity and specificity in fMRI. *NeuroImage*, *172*, 538–553. https://doi.org/10.1016/j.neuroimage.2018.01.078

Sauder, C. L., Derbidge, C. M., & Beauchaine, T. P. (2016). Neural responses to monetary incentives among self-injuring adolescent girls. *Development and Psychopathology*, *28*(1), 277–291. https://doi.org/10.1017/S0954579415000449

Schäfer, T., & Schwarz, M. A. (2019). The Meaningfulness of Effect Sizes in Psychological Research: Differences Between Sub-Disciplines and the Impact of Potential Biases. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00813

Schouppe, N., Demanet, J., Boehler, C. N., Ridderinkhof, K. R., & Notebaert, W. (2014). The role of the striatum in effort-based decision-making in the absence of reward. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(6), 2148–2154. https://doi.org/10.1523/JNEUROSCI.1214-13.2014

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P. J., Dalley, J. W., Flor, H., Gallinat, J., Garavan, H., Heinz, A., Itterman, B., Lathrop, M., Mallik, C., Mann, K., Martinot, J.-L., Paus, T., Poline, J.-B., … IMAGEN consortium. (2010). The IMAGEN study: Reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry*, *15*(12), 1128–1139. https://doi.org/10.1038/mp.2010.4

Schwartz, K. T. G., Kryza-Lacombe, M., Liuzzi, M. T., Weersing, V. R., & Wiggins, J. L. (2019). Social and Non-social Reward: A Preliminary Examination of Clinical Improvement and Neural Reactivity in Adolescents Treated With Behavioral Therapy for Anxiety and Depression. *Frontiers in Behavioral Neuroscience*, *13*. https://doi.org/10.3389/fnbeh.2019.00177

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Steinberg, L., Sharp, C., Stanford, M. S., & Tharp, A. T. (2013). New tricks for an old measure: The development of the Barratt Impulsiveness Scale-Brief (BIS-Brief). *Psychological Assessment*, *25*(1), 216–226. https://doi.org/10.1037/a0030550

Stevens, M. C., Pearlson, G. D., Calhoun, V. D., & Bessette, K. L. (2018). Functional Neuroimaging Evidence for Distinct Neurobiological Pathways in Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *3*(8), 675–685. https://doi.org/10.1016/j.bpsc.2017.09.005

Swartz, J. R., Weissman, D. G., Ferrer, E., Beard, S. J., Fassbender, C., Robins, R. W., Hastings, P. D., & Guyer, A. E. (2019). Reward-Related Brain Activity Prospectively Predicts Increases in Alcohol Use in Adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*. https://doi.org/10.1016/j.jaac.2019.05.022

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, *15*, 29. https://doi.org/10.1186/s12880-015-0068-x

Thompson, W. H., Wright, J., & Bissett, P. G. (2020). Open exploration. *ELife*, *9*, e52157. https://doi.org/10.7554/eLife.52157

Tulloss, R. E. (1997). *Assessment of Similarity Indices for Undesirable Properties and a new Tripartite Similarity Index Based on Cost Functions*.

Urošević, S., Luciana, M., Jensen, J. B., Youngstrom, E. A., & Thomas, K. M. (2016). Age associations with neural processing of reward anticipation in adolescents with bipolar disorders. *NeuroImage. Clinical*, *11*, 476–485. https://doi.org/10.1016/j.nicl.2016.03.013

van den Bergh, D., Haaf, J. M., Ly, A., Rouder, J. N., & Wagenmakers, E. (2019). *A Cautionary Note on Estimating Effect Size.* https://doi.org/10.31234/osf.io/h6pr8

Vasishth, S., Mertzen, D., Jäger, L. A., & Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *Journal of Memory and Language*, *103*, 151–175. https://doi.org/10.1016/j.jml.2018.07.004

Vassena, E., Deraeve, J., & Alexander, W. H. (2020). Surprise, value and control in anterior cingulate cortex during speeded decision-making. *Nature Human Behaviour*, 1–11. https://doi.org/10.1038/s41562-019-0801-5

Veroude, K., von Rhein, D., Chauvin, R. J. M., van Dongen, E. V., Mennes, M. J. J., Franke, B., Heslenfeld, D. J., Oosterlaan, J., Hartman, C. A., Hoekstra, P. J., Glennon, J. C., & Buitelaar, J. K. (2016). The link between callous-unemotional traits and neural mechanisms of reward processing: An fMRI study. *Psychiatry Research. Neuroimaging*, *255*, 75–80. https://doi.org/10.1016/j.pscychresns.2016.08.005

von Rhein, D., Cools, R., Zwiers, M. P., van der Schaaf, M., Franke, B., Luman, M., Oosterlaan, J., Heslenfeld, D. J., Hoekstra, P. J., Hartman, C. A., Faraone, S. V., van Rooij, D., van Dongen, E. V., Lojowska, M., Mennes, M., & Buitelaar, J. (2015). Increased neural

responses to reward in adolescents and young adults with attention-deficit/hyperactivity disorder and their unaffected siblings. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(5), 394–402. https://doi.org/10.1016/j.jaac.2015.02.012

Wilson, R. P., Colizzi, M., Bossong, M. G., Allen, P., Kempton, M., Abe, N., Barros-Loscertales, A. R., Bayer, J., Beck, A., Bjork, J., Boecker, R., Bustamante, J. C., Choi, J. S., Delmonte, S., Dillon, D., Figee, M., Garavan, H., Hagele, C., Hermans, E. J., … MTAC. (2018). The Neural Substrate of Reward Anticipation in Health: A Meta-Analysis of fMRI Findings in the Monetary Incentive Delay Task. *Neuropsychology Review*, *28*(4), 496–506. https://doi.org/10.1007/s11065-018-9385-5

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*, *14*(6), 1370–1386. https://doi.org/10.1006/nimg.2001.0931

Wu, C. C., Samanez-Larkin, G. R., Katovich, K., & Knutson, B. (2014). Affective traits link to reliable neural markers of incentive anticipation. *NeuroImage*, *84*, 279–289. https://doi.org/10.1016/j.neuroimage.2013.08.055

Xu, B., Jia, T., Macare, C., Banaschewski, T., Bokde, A. L. W., Bromberg, U., Büchel, C., Cattrell, A., Conrod, P. J., Flor, H., Frouin, V., Gallinat, J., Garavan, H., Gowland, P., Heinz, A., Ittermann, B., Martinot, J.-L., Paillère Martinot, M.-L., Nees, F., … IMAGEN Consortium. (2017). Impact of a Common Genetic Variation Associated With Putamen Volume on Neural Mechanisms of Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(5), 436-444.e4. https://doi.org/10.1016/j.jaac.2017.02.009

Yarkoni, T. (2009). Big Correlations in Little Studies: Inflated fMRI Correlations Reflect Low Statistical Power-Commentary on Vul et al. (2009). *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *4*(3), 294–298. https://doi.org/10.1111/j.1745-6924.2009.01127.x

Yip, S. W., Kiluk, B., & Scheinost, D. (2019). Toward Addiction Prediction: An Overview of Cross-Validated Predictive Modeling Findings and Considerations for Future Neuroimaging Research. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2019.11.001

Zhang, Y., Padmanabhan, A., Gross, J. J., & Menon, V. (2019). Development of Human Emotion Circuits Investigated Using a Big-Data Analytic Approach: Stability, Reliability, and Robustness. *Journal of Neuroscience*, *39*(36), 7155–7172. https://doi.org/10.1523/JNEUROSCI.0220-19.2019