Interactions between methodological and interindividual variability: How Monetary Incentive Delay (MID) task contrast maps vary and impact associations with behavior

Michael I. Demidenko1 | Alexander S. Weigard1,2 | Karthikeyan Ganesan1 | Hyesue Jang1 | Andrew Jahn3 | Edward D. Huntley4 | Daniel P. Keating1,4

Abstract

Introduction: Phenomena related to reward responsiveness have been extensively studied in their associations with substance use and socioemotional functioning. One important task in this literature is the Monetary Incentive Delay (MID) task. By cueing and delivering performance-contingent reward, the MID task has been demonstrated to elicit robust activation of neural circuits involved in different phases of reward responsiveness. However, systematic evaluations of common MID task contrasts have been limited to between-study comparisons of group-level activation maps, limiting their ability to directly evaluate how researchers’ choice of contrasts impacts conclusions about individual differences in reward responsiveness or brain-behavior associations.

Methods: In a sample of 104 participants (Age Mean = 19.3, SD = 1.3), we evaluate similarities and differences between contrasts in: group- and individual-level activation maps using Jaccard’s similarity index, region of interest (ROI) mean signal intensities using Pearson’s r, and associations between ROI mean signal intensity and psychological measures using Bayesian correlation.

Results: Our findings demonstrate more similarities than differences between win and loss cues during the anticipation contrast, dissimilarity between some win anticipation contrasts, an apparent deactivation effect in the outcome phase, likely stemming from the blood oxygen level-dependent undershoot, and behavioral associations that are less robust than previously reported.

Conclusion: Consistent with recent empirical findings, this work has practical implications for helping researchers interpret prior MID studies and make more informed a priori decisions about how their contrast choices may modify results.

KEYWORDS
Approach, Avoidance, fMRI, Measurement, Monetary Incentive Delay, Prediction Error, Reward Processing
1 | INTRODUCTION

1.1 | Purpose

Due to the hypothesized role of reward systems 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 substrates of approach and avoidance mechanisms during reward processing (Knutson et al., 2000). Univariate contrasts (e.g., Big Win versus Neutral anticipation) that index neural activation during different stages of the MID task have been widely employed to study 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-scale longitudinal studies to index developmental changes in reward mechanisms and their links with negative behavioral outcomes (Case y et al., 2018; Schumann et al., 2010). Despite frequent use of univariate contrasts from this task, there are relatively few studies that have examined how methodological choices made by investigators (e.g., researcher degrees of freedom), such as contrast choice, may impact the underlying results and interpretations about their findings. Therefore, this study aims to clarify the interaction between methodological and interindividual variability in MID task contrast maps and how these interactions affect their associations with psychological measures including substance use and socioemotional functioning.

1.2 | The MID task and theories of reward processing

As of this publication, the MID task has been used in functional magnetic resonance imaging (fMRI) research for almost 20 years and is considered a robust measure of incentive motivation (Knutson & Greer, 2008; Knutson et al., 2000). As an instrumental-reward task, the MID delivers rewards that are contingent on performance involving a timed button response (Richards et al., 2013), whereby different neural regions are recruited depending on whether the reward is being anticipated (i.e., wanted) or consumed (i.e., liked) (Haber & Knutson, 2010). The task was designed to localize reward-related brain activation in substance use populations (Knutson & Heinz, 2015) and identify correlates of individual differences in positive and negative arousal (Wu et al., 2014). A central assumption of the task, inspired in part by the literature on Pavlovian conditioning (Pavlov, 1927) and dopamine responses to positive cues (Schultz, 1998), is that there are brain regions responsible for anticipating and responding to salient stimuli that have positive or negative valence. 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., 2000). The task allows a comparison of valence (positive valence, such as winning, or negative valence, such as losing, across big, or small rewards) and temporal phase (anticipation or outcome).

Activation patterns within anticipation and outcome 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 evoke 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). However, during negative arousal (e.g., loss cue) the MID would elicit avoidance behavior which is reflected by activation in the insula (Knutson & Greer, 2008). Conversely, when modeling the outcome phase (or liking), one would expect less activation of VS (as only ~10% of neurons in nucleus accumbens facilitate pleasure) in response to the pleasure of reward. Hedonic “hot spots” are more likely to be represented in the insula and OFC (Berridge & Kringelbach, 2015) which are reported to be modulated by opioid receptors (Berridge et al., 2010; Buchel et al., 2018; Korb et al., 2020).

It is notable that the specific univariate contrasts used to index reward-related psychological constructs often vary considerably between studies (see Supplemental Table S2). In cases of wanting rewards, reward anticipation is operationalized using contrasts such as All Win versus Neutral (Bourque et al., 2017; Martz et al., 2018; Xu et al., 2017), Big Win versus Neutral (Cao et al., 2019; Cope et al., 2019; Papanastasiou et al., 2018), or Big Win versus Small Win cues (van Hulst et al., 2015; Martz et al., 2016; Stevens et al., 2018). Likewise, in the case of reward consumption, reward outcome is operationalized using contrasts such as Reward Hit versus Neutral Hit (Chan et al., 2016; Mikita et al., 2016; Swartz et al., 2019) or Reward Hit versus Reward Miss cues (Mikita et al., 2016; Navas et al., 2018; Richards et al., 2016). The use of different contrasts to probe the same reward-related constructs is one major source of variability in the MID literature.

The vast majority of fMRI analyses using the MID task focus on specific, unmodulated phases of the task. However, 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 in standard subtraction analysis (Bjork et al., 2010; Oldham et al., 2018). Although reinforcement learning models have been successfully applied to the MID task (Cao et al., 2019), the utility of prediction error still debated (Berridge & O’Doherty, 2014) and it remains to be seen how expected value and prediction error model parameters (positive or negative) modulate the signal in the anticipation and outcome phases during 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). As contingencies in the MID are based on performance, and therefore
relatively uncertain, the MID differs from traditional reinforce-
ment learning paradigms used to investigate prediction errors as the
expectancies are less reliable. Therefore, the MID task may be
considered a proxy to a true temporal-difference learning task
that engenders more reliable expectancies. Nonetheless, previous
work has recommended the use of modulators in the MID task
(Bjork et al., 2010; Oldham et al., 2018), and recent studies have
found that prediction error was positively related to activation in
the bilateral VS (Cao et al., 2019) and substance use problems in
young adults (Cao et al., 2020) during the MID.

**1.3 | Differential use and researcher degrees of freedom in mid task**

Although the MID task has been used extensively to study dys-
functional reward processing in populations with substance use
 disorders (Balodis & Potenza, 2015), it has also been incorporated
into other studies of neurodevelopment and broader psychopathol-
yogy. Various versions of the MID task have been used to investi-
gate reward-related changes as a function of age (Bjork et al., 2010;
Dhingra et al., 2019; Heitzeg et al., 2014), social versus nonsocial
rewards (Schwartz et al., 2019), psychosocial characteristics of im-
pulsivity 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;
Heitzeg et al., 2014; Karoly et al., 2015; Nestor et al., 2019; Sauder
et al., 2016; Schwartz et al., 2019), depression (Chan et al., 2016; Colich
et al., 2017; Landes et al., 2018; Mori et al., 2016), and other psychi-
atric symptoms (Bourque et al., 2017; Lancaster et al., 2016; Maresh
et al., 2019; Mikita et al., 2016; Papanastasiou et al., 2018; von Rhein
et al., 2015; Stevens et al., 2018; Urošević et al., 2016; Veroude
et al., 2016; Xu et al., 2017). Across these studies, a wide range of
brain-behavior effects are reported. In addition to using different
versions of the MID task, the studies cited above often used differ-
ent univariate contrasts to derive activation maps. This raises the
question: "To what extent do analytic methods, such as variation in
univariate contrast selection, inform differences and/or similarities
in conclusions about psychological characteristics?".

Empirical evidence suggests that analytic decisions may result in
substantially different interpretations of fMRI analyses. Carp (2012)
demonstrated that the analytic flexibility in fMRI can generate thou-
sands of statistical maps that can be used in subsequent analyses. As
shown by Botvinik-Nezer et al. (2020), the level of flexibility in task-
based fMRI analyses can produce different outcomes even when
researchers start with identical data and hypotheses. Specifically,
seventy different teams analyzed identical fMRI data with pre-
defined hypotheses regarding risky decision-making. Despite the
similarities across data and hypotheses, between laboratory dif-
fences in contrast selection and region of interest specification
altered the interpretation of results. Thus, without a clear under-
standing of how analytic decisions impact our results and interpre-
tations, 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, it is not well understood how investigators’ ana-
lytic choice of contrasts (e.g., defining anticipation of reward as fol-
lows: $5 Win Cue versus Neutral Cue, or both Win Cues [$5 & $0.20]
versus Neutral Cue) may impact their inferences about the associa-
tion between the neural response to reward and behavior. FMRI ac-
tivation maps differ as a function of reward type/magnitude (Bartra
et al., 2013; Bjork et al., 2010), and recent reviews suggest there is
substantial variability across studies in the techniques used to de-
rive such maps (Balodis & Potenza, 2015; Dugré et al., 2018; Oldham
et al., 2018). Contrast selection is important to the interpretation of
the reported effects because experimental and baseline conditions
are hypothesized to reveal components of a cognitive process which
are reflected in the neural activation (Caplan, 2007). Yet, different
reward contrasts, such as Big Win versus Neutral or Big Win ver-
sus Small Win cues, may be used interchangeably to operationalize
reward anticipation. Combined with publication biases, the diverse
sets of analyses may contribute to underreported contrasts and as-
asociations with behavior that may relate to the arbitrary decisions in
the analytic pipeline (Simmons et al., 2011). Therefore, it is important
to quantify how univariate contrast-related variation in activation
maps within a given sample influences the relative utility of these
maps for predicting behavioral outcomes. This would demonstrate
whether there is a) stability within estimates of activation at each
phase of the task (anticipation or outcome); b) consistency between
conceptually related contrasts in the level of activation in specific
regions of interest (ROI; such that there is higher correlation within
win than between win and loss anticipation); and c) whether choice
between contrasts that, in theory, probe a shared cognitive process,
such as anticipating rewards, alters associations between neural ac-
tivation, and a psychological characteristic.

This would be difficult to deduce from a meta-analysis for sev-
eral reasons. First, meta-analyses typically assess spatial overlap
between contrasts and/or assess relations between different con-
trast activations and external covariates (e.g., behavioral scales or
clinical disorders), but do not assess whether activations from these
contrasts represent distinct versus largely overlapping individual
difference dimensions. Second, most empirical studies report a con-
strained number of MID contrasts, while in some cases making post
hoc justifications for why a particular contrast, or set of contrasts,
was included in the paper. Hence, conclusions from meta-analyses
obfuscate the influence of researcher degrees of freedom linked to
contrast choice and selective reporting.

**1.4 | Current study**

Previous reviews of the MID task have evaluated general utili-
ization of the task in studies of reward responsiveness (Lutz &
Widmer, 2014), between-study, temporal, and phase-related dif-
fferences in MID activation effects (Oldham et al., 2018), dynam-
ics of reward versus loss (Dugré et al., 2018), and influences of
substance use (Balodis & Potenza, 2015) and psychosis profiles (Radua et al., 2015) on activation differences. However, the extent to which contrast choice contributes to variability in activation maps, impacts the measurement of behaviorally relevant individual difference dimensions, 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 univariate contrast 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 to the a) prominent role of motivation (or anticipation of reward) in this task; b) the critical role of dopamine in anticipation (“wanting”) and not outcome (“liking”) (Berridge & Kringelbach, 2015); c) difficulty to temporally differentiate the outcome phase (Bjork et al., 2010); and d) the drop in power during the outcome phase as each anticipatory trial is split into “hit” or “miss” trial outcome, 50% of contrasts focus on the anticipation phase of the MID task. These activation maps are thresholded to compare the degree to which statistical maps (from ten contrasts) a) vary within a phase (e.g., anticipation Big Win > Neutral versus Big Loss > Neutral contrasts) and b) vary between phases of the task (e.g., anticipation versus outcome). The degree of variability is assessed at the individual level and group level to quantify the general pattern in overlap of active voxels between two given contrast’s activation maps. 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 index convergent or divergent dimensions of cognitive processing (such as reward anticipation). Finally, Bayesian correlations between these ROI mean signal intensities and self-reported measures are assessed to determine the impact of contrast choice on the association with psychological measures including substance use, psychosocial, and socioemotional functioning.

While meta-analyses have proposed region-specific activations for positive and negative values across fMRI tasks (Bartra et al., 2013), a recent review of the MID yielded overlapping networks across positive and negative values (Oldham et al., 2018). Given the within-sample comparison of contrasts, instead of testing specific hypotheses within a null hypothesis significance test framework in these analyses, similarities and differences are presented as an index of overlap (Jaccard’s similarity coefficient), and statistical association 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 which specific contrasts the hypotheses are based on in future studies. This exploratory analysis can provide inferences about how contrast selection, which typically precedes the reporting of results and increases researcher degrees of freedom, affects the activation maps. 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 pre-register specific hypotheses related to brain-behavior associations because the intended purpose of the study was to use exploratory analyses to provide a holistic overview of how researcher 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 2,017 (Age Mean = 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 of survey data collection. From Phase I of the study, a subsample of 115 adolescents, who were characterized as high and average/low risk, 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. 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 is made available online (https://github.com/demidenm/MIDContrasts).

2.1 | Self-reported psychological measures

2.1.1 | 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.” Substance use items are identical to those used in the annual, national Monitoring the Future surveys (Johnston et al., 2019).
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.

2.1.2 | 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: (1) = rarely/never, (2) = occasionally, (3) = often, and (4) = almost always/always. A mean score was computed (range: 1 – 4), higher scores indicated lower self-reported impulsivity ($\alpha$=0.79). BIS-B items were z-scored and then aggregated by averaging scores across Wave 1 – Wave 3.

2.1.3 | Sensation seeking

The Brief Sensation Seeking Scale (BSSS) is an 8-item self-report measure of sensation seeking (Hoyle et al., 2002) based on a 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: (1) = strongly disagree, (2) = disagree, (3) = neither disagree nor agree, (4) = agree, and (5) = strongly agree. A mean score was computed (range: 1–5), with higher scores indicated higher self-reported sensation seeking ($\alpha$=0.78). BSSS items were z-scored and then aggregated by averaging scores across Wave 1 – Wave 3.

2.1.4 | 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- to 18-year-olds. The YSR includes two broadband scales: internalizing problems (e.g., withdrawn/depressed) and externalizing problems (e.g., attentional deficit/hyperactivity 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 ($\alpha$ = 0.70 - 0.86) and test–retest reliability ($\alpha$ = 0.67 - 0.88). In the present study, Cronbach's alphas of 0.91 and 0.88 were obtained for the internalizing and externalizing scales, respectively. An aggregate score was created from population-standardized z-scores for internalizing and externalizing by averaging scores across Wave 1 – Wave 3.

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. The modified version in this study is currently being employed in the national Adolescent Brain Cognitive Development (ABCD) study to measure the development of adolescent reward processing (Case y et al., 2018). Identical to the task described in Case y et al. (2018), the task in this study consists of three phases: anticipation, probe, and outcome (i.e., feedback). Each trial starts with a cue type (Win $5, Win $0.20, 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 min 42 s 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 (i.e., 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 the one used in the ABCD study) and that used in the IMAGEN sample (Cao et al., 2019) is the IMAGEN study only includes Win and Neutral trials, thus excluding Loss trials. Furthermore, in the IMAGEN, study performance was rewarded with “points” that were exchanged for M&M’s/candy in contrast to a concrete reward for task performance (e.g., money).

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 = 7,000 ms, TE = 2,900 ms, flip angle = 8°, FOV = 25.6 cm, slice thickness = 1 mm, 208 sagittal slices; matrix = 256x256). Before the MID task, a fieldmap was acquired using spin-echo EPI (TR = 7,400 ms, TE = 80 ms, FOV = 21.6 cm, 90x90 matrix) with opposite phase encoding polarity (A→P, P→A). Two functional T2*-weighted blood-oxygen-level-dependent (BOLD) MID runs were acquired in the axial plane following structural and a face task using a multiband EPI sequence (MB factor = 6) of 60 contiguous axial 2.4 mm slices (TR = 800 ms, TE = 30 ms, flip angle = 52°, FOV = 21.6 cm, 90x90 matrix, volumes = 407).

2.4 | fMRI data analyses

FMRI data were reconstructed; realignment and fieldmap correction was applied in SPM12 to each T2* run to recover inhomogeneity of signal in the B0 field, and physiological noise was removed using RETROICOR (Glover et al., 2000). 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 removed,
and (4) field map correction was applied. The following preprocessing steps were performed: (5) registration to high-resolution structural and standard space MNI 152 image using FLIRT using a Full search 12 DOF (Jenkinson et al., 2002; Jenkinson & Smith, 2001), (6) motion correction using MCFLIRT (Jenkinson et al., 2002), (7) non-brain removal using BET (Smith, 2002), (8) spatial smoothing using a Gaussian kernel of FWHM 5 mm, (9) grand-mean intensity normalization of the entire 4D data set by a single multiplicative factor, and (10) high-pass temporal filtering (Gaussian-weighted least-squares straightline fitting, with sigma = 50.0 s).

3 | fMRI ANALYSES

Subjects were excluded from analyses if a subject’s mean frame-wise displacement (FD) values exceeded > 0.9 within any given run (mean FD pre- and post-preprocessing included in Supplementary Section 1.2), all subjects’ mean post-FD was < 0.9. We focused on commonly used contrasts (Table 1) from a recent review (Oldham et al., 2018) and those from our review of studies using the MID (PubMed 2015 – 2019; Supplementary Table S2), such as reward anticipation (such as Big Win ($5) or All Win ($5 & $0.20) versus Neutral anticipation), Win outcome hit (such as $5 versus Neutral hit 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, win or big win conditions). It should be noted that using anticipation versus outcome phase yields 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). However, as noted in the introduction, the MID is not a true RL design but only a proxy. To incorporate these recommendations, the RL modulators included the following: 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, \( \eta \) = learning rate (0.7)). Similar to Cao et al. (2019), we used a RL model trained by reward cues and outcomes (Rescorla & Wagner, 1972):

\[
EV_t = pGain_t \times \text{Cue}_t \\
PE_t = RR_t \times EV_t \\
pGain_{t+1} = pGain_t + \eta \times PE_t / \text{Cue}_t
\]

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 analysis 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,

| 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 (O6) - Out | Big Win (BW; $5) Hit > Neutral (N) Hit (BWH > NH) |
| Contrast 7 (O7) - 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 (PPE) - BW & SM Modulated |
| Contrast 10 (P10) - PE | Negative Prediction Error (NPE) - BL & SL Modulated |

Note: Ant = anticipation; Out = outcome; individual contrasts modeled in FSL, see section 1.3 in Supplementary for list of events modeled in GLM. A = anticipation; O = Outcome; PE = prediction error
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 2.8 in Supplemental Materials). We extract the mean signal for VS and mPFC in the time-series 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 2.5 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 level 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 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, or spatial overlap, between different activation maps of the MID task within individuals and at the group level. This is to provide an easy to interpret index of how similar (or different) activations are across contrast types. Similar to a previous work (Grady et al., 2020), we calculate a percent overlap using Jaccard's similarity index (JSI) (Maitra, 2010) between contrasts. The JSI calculates the number of voxels that overlap across two thresholded statistical maps. One of the major advantages of using the JSI 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 follows:

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

As we used JSI point estimates to evaluate activated voxels across different thresholded 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 (DiCiccio & Efron, 1996). The bootstrapped JSI would provide reliable estimates of the range and shape of the distribution of percent overlap and a physical interpretation of the JSI obtained across all of the subjects. Although the thresholded maps are impacted by power in the design, similarity can be assessed within phases, such as anticipation or outcome, given the number of trials is comparable within each phase (with the exception of the all win contrast).

3.3 | Region of interest and behavioral associations

Central 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 to create 10-mm-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 extracted using fsmeants. 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 z-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, noninformative 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 0.95 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 (Note: We remind the reader to refer to Table 1 for contrast descriptions). However, for the outcome phase (O6 & O7) four subjects were excluded due to underpowered conditions resulting in anomalies in the estimated [first level & second level] statistical maps, resulting in $N = 100$ for the outcome 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,
FIGURE 1  Mean level activation and deactivation maps for A1-A5 & O6-O7, one-sample t test. See Table 1 for details and online collection for unthresholded statistical maps of tens contrasts https://neurovault.org/collections/6210/
accuracy was lower (48%) and more variable during the neutral condition. Mean response times are not reported, as the E-Prime data were not collected for incorrect (“miss”) trials during the MID task.

JSI similarity matrices and activation maps are displayed in Supplementary Figure S4 and Figure 1, respectively. Associations between individual differences in ROI mean-level activation from each contrast are reported at https://osf.io/a5wem/ and in Figure 2 and are selectively reported below for clarity. Correlations between ROI mean signal intensity estimates and behavioral criterion measures are reported in Figure 3 (subset of four regions, five anticipatory contrasts across our five behaviors; full figure reported in Supplement Figure S5, section 2.7) and available at https://osf.io/d9k3v/. There were four notable patterns present in these results: (1) Win and Loss anticipation demonstrate comparable striatal/insula activation and task-negative deactivation (see NeuroVault statistical map: https://neurovault.org/images/359858/); (2) outcome phase contrasts consistently imply deactivation of striatal regions (potentially due to artifact related to signal spill-over); (3) the Big versus Small Win contrast appears less meaningful than, and unrelated to, other anticipation phase contrasts; 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 were extremely variable in activation and relatively weak in their associations with mean ROI activation from other contrasts; therefore,
4.2 | Big Win and Big Loss anticipation engage similar neural regions

The thresholded masks (p < .001) of A2:BW > N and A5:LB > N group maps had Jaccard’s similarity Coefficient of 0.16 (Supplemental Figure S4). This similarity is also apparent in the group-level activation maps, demonstrated by shared patterns of activation (Figure 1). 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/images/359858/) 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 and 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).

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, O6:BWH > NH & O7:BLH > NH, demonstrated a deactivation of the striatal regions. Based on Jaccard’s similarity Coefficient, 0.34, the regions that were deactivated were comparable in O6:BWH > NH and O7:BLH > NH (Figure 1, and Supplemental Figure S4). Although the mean-level deactivation of the striatal region in the O6:BWH > NH contrast was relatively weak (t = −2.68), in the O7: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 O6:BWH > NH and O7: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 O6:BWH > NH & O7:BLH > NH (https://neurovault.org/images/359858/), O6: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 difference analyses of mean signal intensity in the ACC (r = 0.33), mPFC (r = 0.55), and bilateral VS (left r = 0.45; right r = 0.46). Notably, this deactivation is likely to be a function of the spillover from the anticipatory phase given the short interval between anticipation and outcome stimuli, as can be observed in the BOLD signal change in Figure S6.

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 the lowest Jaccard coefficient with other contrasts modeling the anticipation phase, <0.02 (Figure S4). Further, in the group-level activation, compared to A1:W > N, A2:BW > N, and A5:BL > N anticipation contrasts, the A3:BW > SM had the weakest group-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. For example, the mean-level activation between A1:W > N and A3:BW > SM is negligible: ACC (r = 0.15), mPFC (r = 0.05), bilateral insula (left r = 0.07; right r = 0.08), bilateral OFC (left r = 0.02; right r = 0.06), and bilateral VS (left r = 0.06; right r = 0.15). Yet, there is a strong association between A2:BW > N and A3:BW > SM in the ACC (r = 0.64), mPFC (r = 0.65), bilateral insula (left r = 0.63; right r = 0.58), OFC (left r = 0.60; right r = 0.62), and bilateral VS (left r = 0.59; right = 0.66). Despite the similarity discussed between A2:BW > N and A5:BL > N above in Section 4.2, there is a negligible association between ROI’s in A3:BW > SM and A5:BL > N (r = -0.11 to 0.19), which may suggest 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 expected direction (Supplementary Section 2.4, Table S5). More specifically, there was a strong positive association between internalizing and externalizing problems (r = 0.51), sensation seeking and impulsivity (r = 0.44), externalizing and substance use (r = 0.51), and substance use and sensation seeking (r = 0.36) and impulsivity (r = 0.23).

Figure 3 shows a subset of Bayesian correlations between ROI mean signal intensities and behavioral criterion measures (for complete figure, see Supplemental Figure S5). It shows 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 0.95 probability that the r value falls, respectively (full results available at https://osf.io/d9k3v/; complimentary bootstrapped values are provided at https://osf.io/dr5y2/). Although the interpretation of individual associations is complicated by the large number of tests reported in Figure S5, several general patterns are apparent. First, 71% of the most likely r-values fell at or well below the threshold for what is typically considered a “small-sized” effect, |r| = 0.10 (Supplemental Table S6). Similarly, the bulk of 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| > = 0.30), 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, may be due to multiple testing rather than reflecting true associations. 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.

Second, coupled with the small effects, decisions in contrasts can weaken or alter the brain-behavior results and thus the underlying interpretation. For instance, as can be observed in Figure 3 the median r for the association between anticipatory win activation in the ventral striatum and sensation seeking flips from negative to positive between A1:W > N (right, r = -0.10) and A3:BW > SW (right, r = 0.12). This example, and the high degree of variability in median r between ROI and behaviors presented in Figure 3 and Figure S5, indicates that caution should be taken when selecting contrasts as they may invariably change interpretations even in the context of these small effects.

4.6 | Post hoc analyses

In light of prior meta-analytic 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, O6:BWH > NH versus O7:BLH > NH; win anticipation versus win outcome, A2:BW > N versus O6:BWH > NH; and loss anticipation versus loss gain outcome, A5:BL > N versus O7:BLH > NH. We provide these for reference online https://neurovault.org/collection/.

...some text...

Due to recent concerns that some multiband sequences may alter the BOLD signal in subcortical regions (Risk et al., 2018), we include signal-to-noise ratios and plotted time-series from the VS to provide a direct observation of signal for each anticipation condition.

...some text...

5 | DISCUSSION

In this study of the MID task, we performed an evaluation of similarities and differences between commonly used univariate contrasts, focusing on spatial overlap, individual differences in 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 study provides the first detailed within-study comparison of these common MID task contrasts. The findings demonstrate similarity between positively and negatively arousing anticipation cues, apparent deactivation of striatal regions during the outcome phase, dissimilarity between Big Win > Small Win anticipation and other anticipatory contrasts, and relatively weak associations between MID task activations and self-reported behaviors. These findings are generally consistent with previously reported MID task-specific conceptual findings (Bjork et al., 2010; Oldham et al., 2018) and also have implications for task-general theoretical problems (Hedge et al., 2018; Price & Friston, 1997).

A relatively similar pattern of group-level activation was observed during the Big Win anticipation and the Big Loss anticipation phase. A direct comparison of Big Win versus Neutral and Big Loss versus Neutral anticipation contrasts revealed negligible differences between the activation in the VS 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 other studies reporting similar activation patterns in young adults (Joseph et al., 2016; Murray et al., 2020) and populations at risk to substance use (Bjork et al., 2008). The similarity in the neural activation to the anticipation of Big Win and Big Loss cues is also consistent with the hypothesis that certain regions may display roughly equivalent activation at the extreme ends of value (Bartra et al., 2013). 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; Schoupe...
et al., 2014), as the VS 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). The overlap between win and loss group-level activation suggests the activation maps are more comparable than different which may correspond to a shared cognitive process (Price & Friston, 2005).

However, there was one notable instance in which our analysis revealed dissimilarity between contrasts during the anticipation phase. Although the Big Win versus Small Win contrast activated striatal regions, the contrast demonstrated a limited association with other contrasts in the anticipation phase. Specifically, in group-level activation, there was much greater similarity between Big Win versus Neutral and Big Loss versus Neutral contrasts than the similarity between Big Win versus Neutral and Big Win versus Small Win contrasts. Given that the MID task activates a broad set of regions involved in effortful initiation and anticipation (Suzuki et al., 2020), subtraction of cues with lower effort and greater variability (e.g., neutral stimuli) from higher effort and lower variability (e.g., Big Win), versus with those with slightly more effort (e.g., Small Win), may change the amount of preparatory signal subtracted from the contrast map. It is likely that beyond the cognitive process of “wanting,” there are co-occurring cognitive processes in these cues which may violate assumptions when using subtraction to infer reward sensitivities (Caplan, 2007).

Our comparison of positively and negatively valenced reward outcome contrasts 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 increased activation in reward outcome (Oldham et al., 2018). Oldham et al. (2018) reported increased activation during the outcome phase in the Reward Hit versus Reward Miss or Reward Hit versus Neutral contrasts (see Table 2 in Oldham et al., pg 3404). However, our deactivation results differed from Oldham et al. (2018) in that we focused on the Reward Hit versus Neutral Hit outcome contrast. The observed deactivation of the Reward Hit versus Neutral Hit contrast during the outcome phase is likely the spillover BOLD signal from the anticipatory phase which captures the undershoot (Buxton, 2012). In direct plots of BOLD of outcome within-condition (e.g., Big Win hit and Big Win miss signal), this undershoot is still apparent. Although comparing within-condition outcomes, or more complicated contrasts (Björk et al., 2011; Veroude et al., 2016), are more appropriate when modeling the outcome phase, researchers should remain cognizant that these trials are still unbalanced (e.g., more hit versus miss trials) and underpowered (anticipation trial is bifurcated during outcome). Given the undershoot, if the neural process of interest is specific to the outcome phase, designs that temporally separate the outcome phase should be considered (Björk et al., 2010; Murray et al., 2020).

Bearing in mind that our sample is at the developmental peak of sensation seeking (Romer, 2010; Steinberg et al., 2018), a psychological characteristic that is hypothesized to be central to the motivation toward reward (Casey, 2015; Ernst & Luciana, 2015; Spear, 2011), it is worth to consider how the association between reward activation and sensation seeking changes across anticipatory contrasts. While we found a negligible association between sensation seeking and bilateral VS activation during Big Win versus Neutral contrast (r < -.03), Big Loss versus Neutral has a notable negative association with sensation seeking (r = -.09 - -.10). Then, in the context of the right VS, activation during Big Win versus Small Win contrast and sensation seeking are positively associated (r = .12). These effects may in part be consistent with the hypothesis that higher sensation seekers would be more motivated by positive rewards (e.g., win) and less affected by negative rewards (e.g., loss). However, while these distinctions may be well reasoned from a neurodevelopmental perspective (Casey, 2015) and other work reporting neural associations with sensation seeking (Cservenka et al., 2013; Hawes et al., 2017; Tapia León et al., 2019), the similarity in the negative association between right VS activity and sensation seeking across the All Win versus Neutral (r = -.10) and Big Loss versus Neutral (r = -.09) makes it difficult to discern what the key distinguishing factor is in this brain-behavior association. Although the aforementioned examples refer to the most probable r-values, it is important to remember that the 95% confidence interval in all cases crossed zero and so in some samples the association may include results in the opposite observed direction, which should limit our confidence in the interpretation.

Hence, it is critical to consider how patterns of activation across task phases/conditions relate to behaviors, since the MID task is used in a broad clinical and behavioral literature. In our analysis using psychosocial and clinical criterion measures, we found limited evidence for associations with activations across different phases and conditions. Specifically, the majority of associations between neural activation during the MID task and behavior were likely to be relatively small or negligible. As the original task design focused on clinical populations (Knutson & Heinz, 2015) and reviews suggest a robust role of limbic regions in substance use (Balodis & Potenza, 2015) and psychosis (Radua et al., 2015), this may in part explain the weak effects found in our young adult community sample. Although we cannot rule out that this lack of robust associations with behavior may have been due to features of our sample or measures, it stands in stark contrast to the large array of previous studies reporting associations of MID task activations with various real-world outcomes (Boecker et al., 2014; Büchel et al., 2017). Further, our findings are broadly consistent with recent work that has reported a distinct contrast between the effects found in studies with and without preregistration (median r = 0.16 versus 0.36; Schäfer & Schwarz, 2019) and with findings in large, diverse data sets which indicate that neuroimaging markers often explain only very small portions of the variance in behavioral outcomes of interest (Marek et al., 2020; Nees et al., 2012; Paulus et al., 2019; Paulus & Thompson, 2019). This has led some to suggest that small effects are the “new normal” in clinical neuroscience research (Paulus & Thompson, 2019) and that MRI studies require especially large sample sizes (>2000) to identify meaningful effects in brain-behavior associations (Marek et al., 2020). However, this issue needs to be explored further, as some proposed sample sizes of > 160 in univariate fMRI analyses to be reasonable (Grady et al., 2020).
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. Some studies have reported relatively moderate to large effect sizes ($r > 0.25$) with respect to brain-behavior associations (Cope et al., 2019; Karoly et al., 2015), but despite the numerous brain-behavior tests performed here that focused on related behavioral constructs, our effect sizes were consistently substantially lower (97% out of 400 observations, $r < 0.20$). Until recently, neuroimaging studies of individual differences have frequently been underpowered (Cremers et al., 2017; Yarkoni, 2009), with a median sample size of $< 50$ (Szucs & Ioannidis, 2020), 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). 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. 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 laboratories (Beltz & Weigard, 2019; Paulus & Thompson, 2019) and would be valuable in reexamining the results presented here to understand how effects change.

Beyond the possibility that effect sizes in previous MID studies may have been inflated by small sample sizes and flexible selection of contrasts, the lack of relationships may also be attributed to problematic validity of fMRI-based tasks and the underlying assumptions about the cognitive processes involved, 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 process (Price & Friston, 1997). Although the MID task evokes distinct neural processes that are consistent with current conceptualizations of the mesolimbic system (Knutson & Greer, 2008), the classic metric of validity, namely that a test measures the psychological trait that it claims to measure (Cronbach & Meehl, 1955; Kelley, 1927), appears to be underexplored in the implementation of this paradigm for assessing brain-behavior relationships. In fMRI studies of individual variation, such as behavioral differences that may be associated with neural measures of reward, the combination of experimental and correlational methods is required, work that arises from two distinct traditions in psychology (Cronbach, 1957). Correlation research attempts to increase between-individual variation, whereas experimental work attempts to limit or control for the between-individual variation; the latter methodological approach practice has been argued to contribute to poor predictive effect of cognitive measures in correlational research (Dang et al., 2020). Together, the weak predictive effect of select cognitive tasks and poor test-retest of univariate fMRI (Elliott et al., 2020) can contribute to the unreliable estimates of different task contrasts and the interchangeable use of contrasts will inevitably result in playing “20 questions with nature” (Newell, 1973).

The inferential processes in task-based fMRI pose conceptual challenges. It has been argued that the standard approaches in task-based fMRI that utilize the technique of subtracting conditions are fundamentally flawed in achieving the isolation of the neural substrates of specific mental functions (for discussion, see: Cacioppo et al., 2003; Caplan, 2007; Price & Friston, 2005). Poldrack and Yarkoni (2016) suggest that there are basic conceptual difficulties within subtraction applied in task-based fMRI “that remain widely underappreciated within the neuroimaging community” (pg. 589). This is observed in the MID task, as conceptually the subtraction intends to measure approach and avoidance of positive and negative conditions (Knutson & Greer, 2008), but this is not consistent in the activation patterns of valence (insula) and approach (VS/Nucleus Accumbens) structures that, at the group level, are activated similarly in both conditions (Murray et al., 2020; Oldham et al., 2018). Although using monetary value allows control of magnitude, probability, and timing (Knutson & Greer, 2008), adding a discrete step with positive or negative monetary cues (i.e., “pure insertion assumption”; Price & Friston, 1997) may not be sophisticated enough to identify valence and approach over and above processes of attention and/or motivation within an individual. While the MID task measures distinct positive and negative valenced systems in two distinct phases, the nature to which these phenomena vary or are consistent across specific behaviors has not been well characterized. 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.

Although our findings suggest a high level of variability between contrast choices and behavioral associations, several measures can be taken that may improve the generalizability of results in the MID task literature. First, an immediate step that can be taken by researchers is increasing sample sizes in task-based fMRI research. Currently, a large proportion of fMRI studies are substantially underpowered for finding the effect they are testing (Szucs & Ioannidis, 2017, 2020). Second, researchers would benefit from assessing how the MID contrast values fit in a larger normological network of neural and behavioral constructs (Poldrack & Yarkoni, 2016), beyond an abstract subtraction processes that presume a process of motivation or consumption of reward and preregister these hypotheses in advanced. Third, multivariate methods, such as dimensionality reduction and cross-validated predictive modeling, may help with the reproducibility of theorized neural substrates of cognitive processes (Hong et al., 2019). Multivariate, cross-validated analyses can provide a priori activation patterns and locations that can be confirmed out of sample, reducing the possibility of exploring multiple hypotheses. Finally, if the goal is to characterize individual variability in neural function, researchers should implement functional organization techniques to explain changes in behavior and cognitive processes (Beltz et al., 2016; Yip et al., 2019; Zhang et al., 2019). Network models of task-based fMRI may be particularly helpful for uncovering the neural architecture of cognitive processes (Greene...
et al., 2018; Medaglia et al., 2015). 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 tested only in 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 converge between samples and which are limited to a sample. Future work should examine these associations in a larger sample and at different developmental stages using the ABCD study data. 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 the brain predicts behavior may produce different effects. Moreover, due to a combination of increased number of voxels and alternative methods for controlling the false positive rate, the whole brain statistical analyses exploring brain-behavior associations may reveal findings that an ROI constrained analysis may overlook. 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. 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 the 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, such that increases in activation during Big Win anticipation trials were greater in older adults (Bjork et al., 2010), and reduced activation in response to Big Lose anticipation in 9- to 12-year-olds (Cope et al., 2019). This suggests patterns of activation during the MID task within and between sample comparisons has been considered when age-related effects are present, as qualitative differences between some contrasts may not be easily apparent. Furthermore, whereas these analyses focus on a community-recruited young adult sample, previous reviews focused on clinical population (Balodis & Potenza, 2015; Radua et al., 2015), and these results should be considered in the future within a clinical population to assess how associations would change in light of clinical factors.

6 CONCLUSION

Although univariate fMRI contrasts from the MID task are often used to measure neural substrates of reward processing, modeling techniques have varied substantially between studies. The structure of the task has been proposed to separately measure the constructs of arousal and valence. However, it is still unclear whether these dimensions are easily separable using different task contrasts, and whether findings from different contrasts can be easily generalized between studies. Our within-sample comparison of MID contrasts during multiband fMRI revealed more similarities than differences between positive and negative cues during the anticipation contrast, dissimilarity of the specific Big Win versus Small Win contrast during the anticipation phase, a robust deactivation effect in the outcome phase, and behavioral associations that are less robust than previously thought. These findings point to the need for caution in future work that make attempts at generalization and encourage researchers to power their studies for effects that may be smaller than previously hypothesized.

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 (ST32HD007109-34, V.C. McLoyd & C.S. Monk). A. Weigard was supported by National Institute on Alcohol Abuse and Alcoholism (NIAAA; T32AA007477, F.C. Blow). The authors thank Bennet Fauber, Krishanne Litinas, Christine Wagner, Hani Nasr, Peter Batra, Joshua Hatfield, Meredith House, Kyle Kwaizer, 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 Lelli Mortazavi and Dr. Brian Knutson for sharing code to help facilitate and plot the direct observation of the BOLD signal locked to cue and phase.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR’S CONTRIBUTION

MD conceived the study; MD, KG, and AW conducted the statistical analysis; MD wrote the initial draft of the manuscript, and KG, AW, HJ, and AJ provided support with the analyses, drafts of analyses, and results. DK and EH designed and executed the survey and the neuroimaging protocol. KG, AW, HJ, AJ, EH, and DK assisted MD with study writing and revisions. All authors read and approved the manuscript.

PEER REVIEW

The peer review history for this article is available at https://publon.com/publon/10.1002/brb3.2093.
DATA AVAILABILITY STATEMENT

Readers seeking access to this data should contact the lead author Michael Demidenko (demiddem@umich.edu) or senior author Dr. Daniel Keating (keatingd@umich.edu). Access will be granted to named individuals in accordance with ethical procedures governing the reuse of sensitive data. Infrastructure is currently being developed in collaboration with the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan (https://www.icpsr.umich.edu) to archive and share data for Political and Social Research (ICPSR) at the University of Michigan (https://www.icpsr.umich.edu) to archive and share data in an ethically approved manner and will be shared at a later TBD date.

REFERENCES

Aber, 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). 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. ASEBA. https://trove.nla.gov.au/version/11167240

Aloi, J., Meffert, H., White, S. F., Blair, K. S., Hwang, S., Tyler, P. M., ... Blair, R. J. R. (2019). Differential dysfunctions related to alcohol and cannabis use disorder symptoms in reward and error-processing neurocircuitries 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

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage, 76, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

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 (Berl), 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., Ho, C.-Y., Richard, J. M., & Di Feliceantonio, A. G. (2010). The tempted brain eats: Pleasure and desire circuits in obesity and eating disorders. Brain Research, 1350, 43–64. https://doi.org/10.1016/j.brainres.2010.04.003

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., Knutson, B., & Hommer, D. W. (2008). Incentive-elicited striatal activation in adolescent children of alcoholics. Addiction, 103(8), 1308–1319. https://doi.org/10.1111/j.1360-0443.2008.02250.x

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). Psychosocial problems and recruitment of incentive neurocircuitry: Exploring individual differences in healthy adolescents. Developmental Cognitive Neuroscience, 1(4), 570–577. https://doi.org/10.1016/j.dcn.2011.07.005

Boecker, R., Holz, N. E., Buchmann, A. F., Blomeyer, D., Plichta, M. M., Wolf, I., ... 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., Holmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., ... 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., ... IMAGEN Consortium. (2017). Functional neuro-imaging 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/10.1006/S1053-8119(96)80222-9

Buchel, C., Miedl, S., & Sprenger, C. (2018). Hedonic processing in humans is mediated by an opioidergic mechanism in a mesocorticolimbic system. eLife, 7, e39648. https://doi.org/10.7554/eLife.39648

Büchel, C., Peters, J., Banaschewski, T., Bokde, A. L. W., Bromberg, U., Connord, P. J., ... 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

Buxton, R. B. (2012). Dynamic models of BOLD contrast. NeuroImage, 62(2), 953–961. https://doi.org/10.1016/j.neuroimage.2012.01.012

Cacioppo, J. T., Berntson, G. G., Lorig, T. S., Norris, C. J., Rickett, E., & Nusbaum, H. (2003). Just because you’re imaging the brain doesn’t mean you can stop using your head: A primer and set of first
Dugré, J. R., Dumas, A., Bitar, N., & Potvin, S. (2018). Loss anticipation and outcome during the Monetary Incentive Delay Task: A neuroimaging systematic review and meta-analysis. PeerJ, 6, e4749. https://doi.org/10.7717/peerj.4749

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., ... Hariri, A. R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. Psychological Science, 31(7), 792–806. https://doi.org/10.1177/0956797620916786

Ernst, M. & Luciana, M. (2015). Neuroimaging of the dopamine/reward system in adolescent drug use. CNS Spectrums, 20(4), 427–441. https://doi.org/10.1017/S1092852915000395

Floresco, S. B. (2015). The nucleus accumbens: An interface between cognition, emotion, and action. Annual Review of Psychology, 66, 25–52. https://doi.org/10.1146/annurev-psych-012013-115159

Gelman, A. & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis—A“garden of forking paths”—Explains why many statistically significant comparisons don’t hold up. American Scientist, 102(6), 460–466. https://doi.org/10.1511/2014.111.460

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magnetic Resonance in Medicine, 44(1), 162–167. https://doi.org/10.1002/1522-2594

Gonzalez, M. Z., Allen, J. P., & Coan, J. A. (2016). Lower neighborhood quality in adolescence predicts higher mesolimbic sensitivity to reward anticipation in adulthood. Developmental Cognitive Neuroscience, 22, 48–57. https://doi.org/10.1016/j.dcn.2016.10.003

Grady, C. L., Rieck, J. R., Nichol, D., Rodrigue, K. M., & Kennedy, K. M. (2020). Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fMRI data. Human Brain Mapping, https://doi.org/10.1002/hbm.25217

Greene, A. S., Gao, S., Scheinost, D., & Constable, R. T. (2018). Task-induced brain state manipulation improves prediction of individual traits. Nature Communications, 9(1), 1–13. https://doi.org/10.1038/s41467-018-04920-3

Haber, S. N. & Knutson, B. (2010). The Reward circuit: Linking primate anatomy and human imaging. Neuropsychopharmacology, 35(1), 4–26. https://doi.org/10.1038/nn.2009.129

Hayler, D. J., Hatton, S. N., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., ... 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

Hawes, S. W., Chahal, R., Hallquist, M. N., Paulsen, D. J., Geier, C. F., & Luna, B. (2017). Modulation of reward-related neural activation on sensation seeking across development. Neuroimage, 147, 763–771. https://doi.org/10.1016/j.neuroimage.2016.12.020

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

Heitzeg, M. M., Villafuerte, S., Weiland, B. J., Enoch, M.-A., Burmeister, M., Zubieta, J.-K., & Zucker, R. A. (2014). Effect of GABRA2 genotype on development of incentive-motivation circuitry in a sample enriched for alcoholism risk. Neuropsychopharmacology, 39(13), 3077–3086. https://doi.org/10.1038/nn.2014.161

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., Palmgren, 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
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

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., ... Dosenbach, N. U. F. (2020). Towards reproducible brain-wide association studies. *BioRxiv*, 2020.08.21.257758. https://doi.org/10.1101/2020.08.21.257758

Mares, 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

Martz, M. E., Zucker, R. A., Schulenberg, J. E., & Heitzeg, M. M. (2018). Psychosocial and neural indicators of resilience among youth with a family history of substance use disorder. *Drug and Alcohol Dependence*, 185, 198–206. https://doi.org/10.1016/j.drugdep.2017.12.015

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., ... 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., ... Yamawaki, S. (2016). Behavioral activation can normalize neural hyperactivity 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

Murray, L., Lopez-Duran, N. L., Mitchell, C., Monk, C. S., & Hyde, L. W. (2020). Neural mechanisms of reward and loss processing in a low-income sample of at-risk adolescents. *Social Cognitive and Affective Neuroscience*, 15(12), 1310–1325. https://doi.org/10.1093/scan/nswa157

Navas, J. F., Barrós-Loscertales, A., Costumero-Ramos, V., Verdejo-Román, J., Vilar-López, R., & Verdejo-García, A. (2018). Excessive body fat linked to blunted somatosensory cortex response to general reward in adolescents. *International Journal of Obesity*, 42(1), 88–94. https://doi.org/10.1038/ijo.2017.207

Nees, F., Tschoppe, J., Patrick, C. J., Vollstädt-Klein, S., Steiner, S., Poustka, L., ... 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

Oldham, S., Murawski, C., Fornito, A., Yousef, 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., … IMAGEN Consortium. (2018). Examination of the neural basis of psychotic-like 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., … 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

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (pp. xv–430). Oxford University Press.

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

Price, C. J. & Friston, K. J. (1997). Cognitive conjunction: A new approach to brain activation experiments. *NeuroImage, 5*(Pt 1), 261–270. https://doi.org/10.1006/nimg.1997.0269

Price, C. J. & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. Cognitive *Neuropsychology, 22*(3), 262–275. https://doi.org/10.1080/0264329044200095

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., Schlagenauf, 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 I. A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton-Century-Crofts.

Richards, J. S., Arias Vásquez, A., von Rhein, D., van der Meer, D., Franke, B., Hoekstra, P. J., … Hartman, C. A. (2016). Adolescent behavioral and neural reward sensitivity: A test of the differential susceptibility theory. *Translational Psychiatry, 6*, e771. https://doi.org/10.1038/tp.2016.37

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

Romer, D. (2010). Adolescent risk taking, impulsivity, and brain development: Implications for prevention. *Developmental Psychobiology, 52*(3), 263–276. https://doi.org/10.1002/dev.20442

Sauder, C. L., Derbridge, 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

Scherer, 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*, 12. https://doi.org/10.3389/fpsyg.2019.00813

Schoupe, N., Demanet, J., Boehler, C. N., Riddinkhof, 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*(16), 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., … Struve, M. (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*, 103389. 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), 143–55.

Spear, L. P. (2011). Rewards, aversions and affect in adolescence: Emerging convergences across laboratory animal and human data. *Developmental Cognitive Neuroscience, 1*(4), 390–403. https://doi.org/10.1016/j.dcn.2011.08.001

Steinberg, L., Icnenogle, G., Shulman, E. P., Breiner, K., Chein, J., Bacchini, D., … Takash, H. M. S. (2018). Around the world, adolescence is a time of heightened sensation seeking and immature self-regulation. *Developmental Science, 21*(2), https://doi.org/10.1111/desc.12532

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*, 83(6), 2148–2154. https://doi.org/10.1016/j.biopsych.2017.09.005

Suzuki, S., Lawlor, V. M., Cooper, J. A., Arulpragasan, A. R., & Treadway, M. T. (2020). Distinct regions of the striatum underlying effort, movement initiation and effort discounting. *Nature Human Behaviour*, 1–11. https://doi.org/10.1038/s41562-020-00972-y

Swartz, J. R., Weissman, D. G., Ferrer, E., Beard, S. J., Fassbender, C., Robins, R. W., … Geyer, 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, 58*(3), 391–400. 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

Szucs, D. & Ioannidis, J. P. A. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage, 221*, 117164. https://doi.org/10.1016/j.neuroimage.2020.117164

Tapia León, I., Kruse, O., Stark, R., & Klucken, T. (2019). Relationship of sensation seeking with the neural correlates of appetitive
conditioning. Social Cognitive and Affective Neuroscience, 14(7), 769–775. https://doi.org/10.1093/scan/nsz046
Thompson, W. H., Wright, J., & Bisselt, P. G. (2020). Open exploration. Elife, 9, e52157. https://doi.org/10.7554/elife.52157
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/h0pr8
van Hulst, B. M., de Zeeuw, P., Lupas, K., Bos, D. J., Noggers, S. F. W., & Durston, S. (2015). Reward anticipation in ventral striatum and individual sensitivity to reward: A pilot study of a child-friendly fMRI task. PLoS One, 10(11), e0142413. https://doi.org/10.1371/journal.pone.0142413
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., ... 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., Zwiets, M. P., van der Schaaf, M., Franke, B., Luman, 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., Kempston, M., Abe, N., ... 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., ... 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(7), 7155–7172. https://doi.org/10.1523/JNEUR0220-19.2019

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Demidenko MI, Weigard AS, Ganesan K, et al. Interactions between methodological and interindividual variability: How Monetary Incentive Delay (MID) task contrast maps vary and impact associations with behavior. Brain Behav. 2021;11:e02093. https://doi.org/10.1002/brb3.2093