



Turning down the heat: Neural mechanisms of cognitive control for inhibiting task-irrelevant emotional information during adolescence

Marie T. Banich^{a,b,*}, Harry R. Smolker^b, Hannah R. Snyder^c, Jarrod A. Lewis-Peacock^d, Detre A. Godinez^e, Tor D. Wager^{a,b}, Benajmin L. Hankin^f

^a Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO 80309, United States

^b Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO 80309, United States

^c Department of Psychology, Brandeis University, Waltham, MA 02453, United States

^d Department of Psychology, University of Texas at Austin, Austin, TX 78712, United States

^e Department of Psychology, University of Denver, Denver, CO 80208, United States

^f Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL 61820, United States

ARTICLE INFO

Keywords:

Cognitive control
fMRI
Multi-voxel pattern analysis
DLPFC
Amygdala
Individual differences

ABSTRACT

One major question in the cognitive neuroscience of cognitive control is whether prefrontal regions contribute to control by upregulating the processing of task-relevant material or by downregulating the processing of task-irrelevant material. Here we take a unique approach to addressing this question by using multi-voxel pattern analysis, which allowed us to determine the degree to which each of the task-relevant and task-irrelevant dimensions of a stimulus are being processed in posterior cortex on a trial-by-trial basis. In our study, adolescent participants performed an emotion word – emotional face Stroop task requiring them to determine the emotional valence (positive, negative) of a task-relevant word in the context of a task-irrelevant emotional face. Using mediation models, we determined whether activation of a major cognitive control region, the dorsolateral prefrontal cortex (DLPFC), influences reaction time on a trial-by-trial basis directly or if it does so indirectly by modulating processing of the task-relevant and/or task-irrelevant information in posterior brain regions. To examine the specificity of the effects observed for the DLPFC, similar analyses were performed for the amygdala, a brain region involved in processing of the salient task-irrelevant emotional information.

For both congruent and incongruent trials, increased DLPFC activity on a given trial was associated with reduced perceptual processing of the task-irrelevant face, consistent with the idea that top-down cognitive control can modulate processing of task-irrelevant information. No effect of DLPFC activity was observed on processing of the task-relevant word. However, increased processing of the task-relevant word was associated with longer RT on congruent trials but not incongruent trials, which may reflect a need for greater processing of the task-relevant word to overcome any influence of the pre-potent task-irrelevant face.

In a more exploratory aspect of our investigation, multi-level moderated mediation models were used to examine the influence of individual differences on the observed brain-behavior relationships. For congruent trials, the influence of task-irrelevant face processing on RT was decreased in individuals with higher self-reported Executive Control and increased in those with higher levels of self-reported Negative Affect.

These results suggest that cognitive control regions in prefrontal cortex during adolescence can suppress the processing of task-irrelevant information in sensory cortex to influence performance (RT). The processing of task-relevant information may also influence performance, but such processing did not reveal evidence of being modulated by cognitive control regions. Moreover, these effects are sensitive to individual differences in the self-reported ability to exert cognitive and affective control. As such, we provide insights into the more precise mechanisms by which cognitive control influences task performance on a trial-by-trial basis during adolescence.

1. Introduction

Most theories of cognitive control posit that prefrontal regions act to

exert control by modulating the processing of posterior brain regions (Miller and Cohen, 2001; Banich, 2009; Banich et al., 2000), and much empirical research supports this assertion (Zanto et al., 2011; Woolgar

* Correspondence to: Institute of Cognitive Science, Muenzinger E230, University of Colorado Boulder, Boulder, CO 80309-0344, United States.

E-mail address: Marie.Banich@colorado.edu (M.T. Banich).

et al., 2011). Cognitive control is required especially when one must pay attention to task-relevant material in the presence of distracting or pre-potent task-irrelevant material. However, one outstanding question regards whether prefrontal modulation of attention acts by up-regulating processing of task-relevant material or by down-regulating task-irrelevant material. Hypothetically, control could be exerted through either one of these mechanisms or by both in tandem. Here we examine that issue in an adolescent sample, in whom the development of the neural mechanisms for cognitive control may still be developing (e.g., Andrews-Hanna et al., 2011). To our knowledge, this issue has only been explored previously in adults.

One task used to address this issue is the Stroop task, and related variations, which are classic measures of cognitive control. In Stroop tasks, individuals are required to pay attention to a task-relevant dimension of a bivalent stimulus, while ignoring another salient, pre-potent or more automatically processed task-irrelevant dimension. For example, in the classic color - word Stroop task, individuals must identify the color ink in which a word is printed. Generally, performance is compared between two conditions: an incongruent condition and a congruent condition. In the incongruent condition, which requires a higher level of cognitive control, the meaning and response generated by the more automatically processed task-irrelevant dimension, a word (e.g., “red”), conflicts with the information contained in the task-relevant dimension, the ink color (e.g., blue). In the less demanding congruent condition, no such conflict exists (e.g., the word “red” displayed in red ink).

Research performed so far with Stroop tasks to address the question of whether task-relevant information is up-regulated, or task-irrelevant information is down-regulated has been equivocal. Early neuroimaging studies using PET examined the degree to which activation changes, on a group level, were observed in brain areas known to process the task-relevant and task-irrelevant dimensions, respectively. The results of these studies were varied with some showing modest evidence that activation in regions likely to process the task-irrelevant word was reduced, suggestive of inhibition (Bench et al., 1993), some showing evidence for both decreased activation in regions likely to process the task-irrelevant word and increased activation in regions likely to be process the task-relevant color (Carter et al., 1995), and some showing no effect (Pardo et al., 1990).

One difficulty in interpreting these studies is that there is significant variation amongst individuals in exactly which portions of the brain support the processing of specific visual stimulus dimensions, such as words or colors. To overcome this issue, subsequent researchers defined the brain regions that process a given stimulus dimension on an individual basis via a localizer task, for example, identifying an individual's fusiform face area. The researchers then examined changes within those regions on an individual basis. The results of these studies have been equivocal as well. While some find that evidence for up-regulation of the task relevant dimension (Egner and Hirsch, 2005; Purmann and Pollmann, 2015), others find evidence that both task-relevant information can be up-regulated and that task-irrelevant information can be down-regulated (Polk et al., 2008; Coste et al., 2011).

Here we address this question of whether control works by modulating the processing of task-relevant information or task-irrelevant information in a number of novel ways. First, we leverage the power of multi-voxel pattern analysis (MVPA) to determine, on each trial, the degree to which participants are separately processing the task-relevant dimension of the stimuli and the task-irrelevant dimension of the stimuli. MVPA is a machine learning technique that provides sensitive readouts of mental representations based on distributed patterns of activation across brain regions (see Lewis-Peacock and Norman, 2014 for review). Much work suggests that activation of the processing of specific visual features is distributed across a wider expanse of visual cortex than classically defined category-specific processing regions (e.g., fusiform face area) (Haxby et al., 2001, 2014). In fact, MVPA can be more sensitive to detecting processing of information than localized

patterns of activation as determined by the standard GLM approach to analyzing neuroimaging data (Lewis-Peacock and Postle, 2012). Hence, we reasoned that MVPA would provide a more sensitive measure of the degree to which each of the task-relevant and task-irrelevant dimensions of the Stroop stimulus are being processed.

Second, we link such an approach to behavior. More specifically, we examine whether the *degree* of activation in each of these brain systems (i.e., that processing the task-relevant dimension, that processing the task-irrelevant dimension) can predict variations in behavior (i.e., RT) via multi-level modeling and mediation (Hayes, 2013) as applied to brain imaging data (e.g., Wager et al., 2009). This approach allowed us to examine the relationship between brain activation and behavior on a trial-by-trial basis while simultaneously taking into account between-subject variability in both brain dynamics and behavior.

Third, we compare how the effects of a cognitive control region on processing task-relevant and task-irrelevant aspects of a Stroop stimulus might differ from the effects exerted by a brain region not involved in cognitive control, but which nonetheless is linked specifically to the processing of the task-irrelevant dimension. Said differently, we wished to examine the degree to which the relationships we observe between activity in cognitive control regions, the degree of processing of each of the stimulus dimensions and reaction time is specific to control regions as compared to another brain region that might influence processing of the task-irrelevant stimulus dimension. To do so, we recorded brain activation while participants performed a variant of the Stroop task requiring them to categorize the affective valence (positive or negative) of a task-relevant word superimposed on a to-be-ignored task-irrelevant emotional face (with either a happy or sad expression). In our task, incongruent trials were those in which the task-relevant word (e.g., “happy”) was superimposed on a task-irrelevant face displaying a conflicting emotional expression (e.g., a sad expression). Congruent trials were those in which the task-relevant word (e.g., “happy”) was superimposed on a task-irrelevant face displaying the same emotional expression. It is well known that reaction times (RTs) are elongated on incongruent trials when cognitive control must be exerted to ignore a conflicting task-irrelevant dimension as compared to congruent trials in which no such conflict is present, indicative of greater control requirements on the former type of trial as compared to the latter. As such we used RT as our behavioral measure of interest. We also investigated brain-behavior relationships for incongruent and congruent trials, separately. This approach allowed us to determine whether modulation of task-relevant or task-irrelevant information varied by the amount of control required.

Our proxy for activation in the cognitive control system was a dorsolateral prefrontal cortex (DLPFC) region of interest (ROI), a region typically implicated in cognitive control (e.g., Niendam et al., 2012). Acting as a comparison region was our proxy for activation in the systems sensitive to the emotional salience of the task-irrelevant face, an amygdalar ROI, encompassing a region that shows heightened activity related to emotional processing (Pourtois et al., 2013) (see Section 2 for how the exact location of these ROIs were selected).

We examined the degree to which the perceptual processing of each of these aspects of the stimulus (task-relevant information, task-irrelevant information), influenced behavior (i.e., RT). Importantly, evidence suggests that cognitive control regions modulate processing of perceptual information depending on its task relevance (e.g. Zanto et al., 2010), while the amygdala affects the processing of perceptual information based on emotional salience (e.g., Murray et al., 2014; Vuilleumier, 2015). Hence, we examined the degree to which these brain systems (i.e., DLPFC, amygdala) exert their effect on behavioral performance by altering processing of each of the task-relevant and task-irrelevant stimulus dimensions.

We chose to examine this issue in an adolescent population both for theoretical reasons and with regards to its real-world relevance. Leading theories of adolescent brain development currently focus on the relationship between regions involved in cognitive control and

those involved in emotion processing (Steinberg, 2010; Shulman et al., 2016; Casey et al., 2016; Ernst, 2014). Brain areas involved in cognitive control, including prefrontal and parietal regions, are structurally and functionally still maturing (e.g., Andrews-Hanna et al., 2011; Mills et al., 2016), while brain regions, such as the amygdala, involved in processing emotional information are particularly active compared to younger children or young adults (Guyer et al., 2016).

As such, focusing our investigation on this age group provides a particular tractable and informative way to compare the influence of top-down control regions (i.e., DLPFC) as compared to the influence of another brain region, in this case the amygdala, known to process the task-irrelevant dimension of our stimuli. While numerous brain imaging studies have examined neural systems supporting cognitive control in adolescents using the Stroop task (e.g., Vijayakumar et al., 2014; Veroude et al., 2013; Andrews-Hanna et al., 2011; Adleman et al., 2002), we know of none that has specifically focused on the degree to which the processing of task-relevant vs. task-irrelevant information is modulated. Note that the purpose of this study was not to directly compare the pattern observed in adolescents to adult participants, which is the focus of an on-going study.

From the vantage point of real-world relevance, our investigation was also designed to examine how individual differences might moderate these brain-behavior relationships. In particular, adolescence is a developmental time period in which aspects of psychopathology first become apparent (e.g., Paus et al., 2008; Hankin et al., 2015). As such, our investigation examined how individual differences predictive of psychopathology might moderate these brain-behavior relationships. In particular, we examined negative affect, a risk factor for psychopathology, and executive control, a protective factor (e.g., Hankin, 2015; Hankin et al., 2016). Prior research has shown associations between such traits and brain function in adolescents: negative affect is associated with increased amygdala activity, especially when viewing sad faces (Henderson et al., 2014) such as were used here, whereas increased self-reported executive control is linked to increased activity of dorsolateral prefrontal cortex (Andrews-Hanna et al., 2011). We employed a Stroop task because the standard Stroop task and its variants has been found sensitive to detecting alterations in brain function and behavior in adolescents as function of individual differences with regards to risk or presence of psychopathology (e.g., Aloï et al., 2018; Banich et al., 2007; Killgore et al., 2007). We posited that individual differences might moderate the observed brain-behavior relationships, and the Stroop task would be a sensitive way to measure them.

It was difficult to develop strong *a priori* hypotheses since control could be exerted through several different pathways, and little prior research with regards to this question in this age group exists. Nonetheless, consistent with standard models of cognitive control, we posited that DLPFC activation would influence either the processing of the task-relevant word or the task-irrelevant face, or both. We further posited that we would observe differences between incongruent and congruent trials given the differences between the two in the amount of control required. Finally, we predicted that the relationships we observed for the DLPFC would be distinct from that observed for the amygdala.

2. Materials and methods

2.1. Participants

A total of 32 healthy, right-handed adolescents (14 female), ages 14–17, (mean = 15.8 years, sd = 0.9 years) were recruited from the greater Denver/Boulder metropolitan area through online advertisements and flyers and were paid for their participation. We centered our sample around age 16 because this is the age at which adolescents tend to have reached adult levels in basic cognitive capacities (e.g., Steinberg et al., 2009), have completed puberty (Costello et al., 2011) and are still in a somewhat homogenous environment (i.e., attending

high school) (Arnett, 2000). Data from four participants were dropped because they had head motion greater than 3.0 mm; this resulted in final data set of 28 participants. In addition, functional localizer data for 3 participants was excluded from the MVPA analyses due to head motion, leading to an *N* of 25 for these analyses. Written informed parental consent and participant assent were obtained prior to the experimental session. All experimental protocols were approved by CU-Boulder's Institutional Review Board prior to data collection.

2.2. Experimental procedure

Participants came in for a 3-h experimental session. After informed consent was obtained, participants completed the neuroimaging portion of the session. Participants first performed two functional localizer tasks, the goal of which was to enable characterization, via MVPA, of the neural signatures of activity in posterior brain regions specific to processing the task-relevant word processing and separately the processing of the task-irrelevant face. Next, individuals performed a spatial word - object Stroop task, which provided the ROIs of interest for the present study. In this task, individuals made a decision based on the spatial meaning of a word (e.g. above or below), while ignoring the position of the word relative to an image on which it was superimposed. This task was followed by the main task of interest, the emotional word - emotional face Stroop task. In this task, individuals made a decision based on the emotional valence of a word (positive or negative), while ignoring the emotional valence of the face on which the word was superimposed. Task order of the Stroop tasks was not counterbalanced to avoid a situation in which the faces with neutral expressions in the spatial word - object Stroop task were imbued with emotional significance as might have occurred if they were viewed following the emotional word - emotional face Stroop task (see Fig. 1). After the neuroimaging portion of the session, participants completed questionnaires to assess self-reported Executive Control and Negative Affect, as well as the 4-subtest version of the Wechsler (1999).

2.3. Stimuli

Stimuli for the two functional localizer tasks consisted of black-and-white photographs of birds, houses, faces, and words. The bird stimuli consisted of a single set of 32 grey-scale images of birds obtained from the bird-identification field guides or ornithological websites, and standardized to control for background and image size. Half of the bird images depicted a bird facing to the left, while the other half depicted a bird facing to the right. House stimuli, used in the localizer and spatial word-object Stroop tasks were drawn from a standardized set of images previously used in studies of visual processing (Ishai et al., 1999, 2000). Sixty-four house images were used across the two localizer tasks (32 images for each task), while 16 images of houses were used in the spatial word - object Stroop task, with each house image presented four times over the course of the task. Across all tasks employing house stimuli, half of the house images showed a house with a deck and half showed a house without a deck.

All face images were drawn from the NimStim Set of Facial Expressions (Tottenham et al., 2009; <http://www.macbrain.org/resources.htm>). Images from 23 different posers (11 female; 12 male) were used across the two localizer tasks while images from 16 posers (8 male; 8 female) were used for the spatial word - object Stroop task and the emotional word - emotional face Stroop tasks. For the localizer tasks, a “happy”, “sad”, and “neutral” facial expression image was selected for each of the 23 posers, resulting in a set of 69 images. Of these 69 images, 64 images were included in the localizer tasks (32 images for each task) and each face was presented once. Images were pseudo-randomly selected to be in one localizer task or the other, ensuring that an equal number of male and female faces were shown within both tasks, balanced across valence (10 happy, 10 sad, 12 neutral). In the spatial word - object task, one image was used from each poser

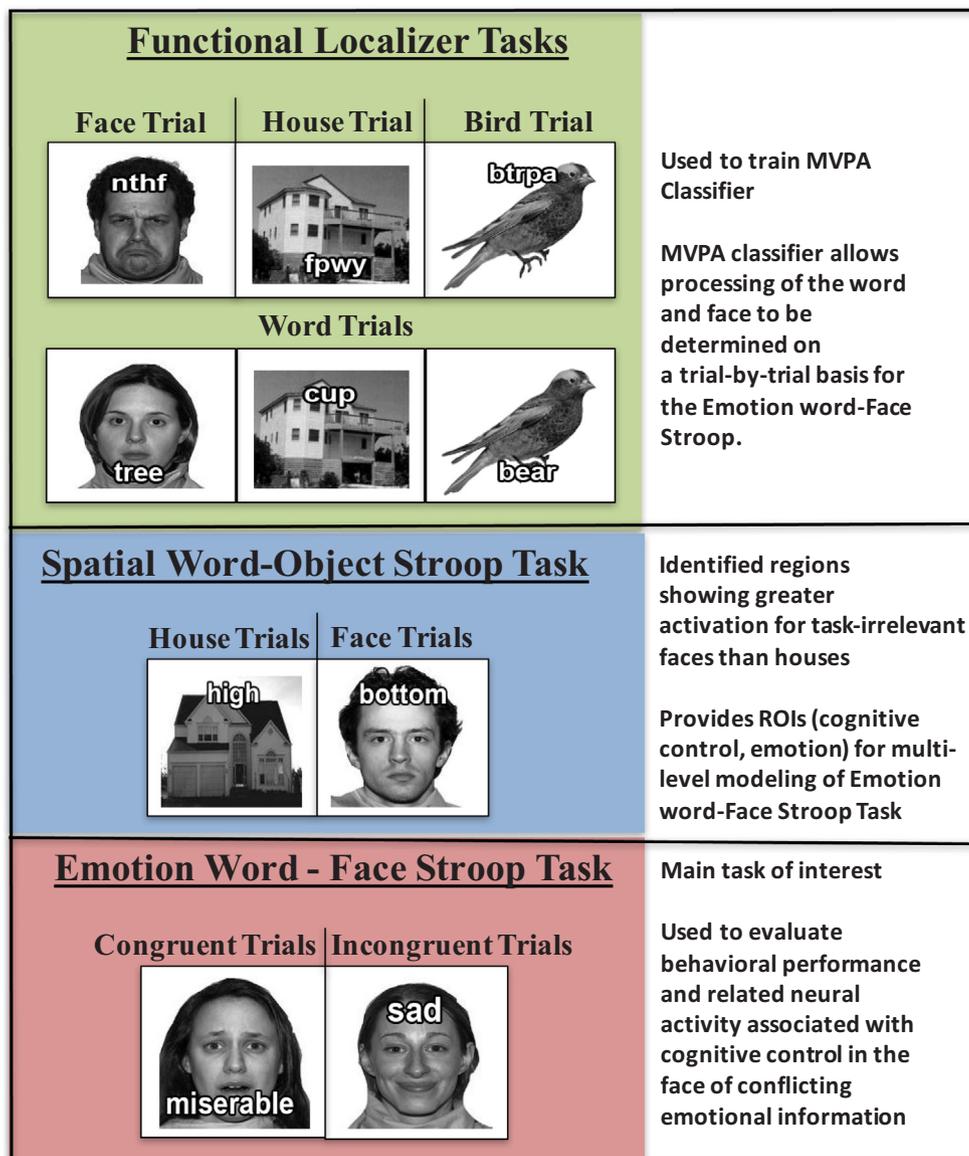


Fig. 1. The tasks used in the current study and the measures derived from each. (top) Individuals were given a localizer task to allow MVPA classification of words and faces, dimensions in the main task of interest. (middle) A non-emotional spatial word-object Stroop task was used to identify brain regions involved in cognitive control and emotion processing, respectively, through identification of those regions that exhibited greater activation when a neutral face as compared to a house was the task-irrelevant dimension. (bottom) The emotional word-face Stroop task designed to assess activation of cognitive control and emotional processing regions when adolescents must ignore task-irrelevant emotional information contained in facial expressions.

displaying a neutral facial expression and each image was shown four times over the course of the task. In the emotional word - emotional face task, two images were used from each poser, one depicting a “happy” facial expression (positive valence) and the other depicting a “sad” facial expression (negative valence), and each image was displayed four times over the course of the task. Across all tasks, words were printed in white ink with a black border in bold Arial font superimposed on a black-and-white picture.

2.4. Questionnaires

Two sets of self-report questionnaires that have been well validated in adolescents, one to assess individual differences in Executive Control (EC) and another to assess Negative Affect (NA), were used. These measures both have been shown to explain significant variance in psychopathology in both adolescents and adults (Tackett et al., 2013; Tellegen et al., 1999). To assess self-reported EC, we used (1) the effortful control scale of the *Early Adolescent Temperament Questionnaire-Revised (EATQ-R)* (Ellis and Rothbart, 2001), which assesses effortful control via three subscales: activation control (e.g., “Has a hard time finishing things.”), attention (e.g., “Finds it easy to really concentrate on a problem”), and inhibitory control (e.g., “Opens presents before s/

he is supposed to.”), (2) the *Attentional Control Scale (ACS)* (Derryberry and Reed, 2002), which assesses voluntary attentional control (e.g., “When I am working hard on something, I still get distracted by events around me.”), and (3) the *Behavioral Rating Inventory of Executive Function Self-Report Version (BRIEF-SR)* (Gioia et al., 2000) which assesses multiple areas of executive functioning via Inhibit, Shift, Emotional Control, Monitor, Working Memory, Plan/Organize, Organization of Materials, and Task Completion subscales. To assess self-reported NA, we used (1) the negative emotional mood scale of the *Early Adolescent Temperament Questionnaire-Revised (EATQ-R)*, (Ellis and Rothbart, 2001), (2) *Children's Depression Inventory (CDI)* (Kovacs, 1992), which is the most commonly used measure of depressive symptoms in adolescents, (3) the *Multidimensional Anxiety Scale for Children (MASC)* (March et al., 1997), which assesses physical anxiety symptoms, social anxiety, harm avoidance, and separation anxiety/panic, and (4) the *Penn State Worry Questionnaire for Children (PSWQ-C)* (Chorpita et al., 1997), which assesses excessive worry (anxious apprehension). A composite score was computed for EC and another for NA by determining the average of each participant's Z-scores (calculated across participants) across the questionnaires relevant for a given construct.

2.5. Neuroimaging

2.5.1. Data acquisition

A SIEMENS MAGNETOM Trio (3-Tesla) MRI system with a 12-channel head coil was used for data acquisition. Acquisition parameters for both the functional tasks were repetition-time [TR] = 2300 ms (Stroop), 2000 ms (functional localizers); echo time [TE] = 25 ms; flip angle = 73°, with each image consisting of 38 contiguous slices (thickness = 3 mm; in-plane resolution = 3 mm), with slices aligned parallel to the orbital frontal cortex. For each localizer task, a total of 130 EPI images were acquired. The spatial word-object and emotional word-emotional face Stroop tasks were each presented in a separate run, with a total of 261 EPI images acquired per run. Foam padding was placed around the head, within the head coil, to limit head motion during the scan.

2.5.2. Tasks

2.5.2.1. Localizer tasks. The goal of the localizer tasks was to train multi-voxel pattern classifiers to determine, for each participant, the particular patterns of activation across posterior brain regions associated with perceiving specific visual stimuli (words, faces, houses, birds). In this manner, the classifiers derived from the localizer tasks could then be used to assess the degree of perceptual processing being afforded each of the stimulus dimensions in our emotional word – emotional face Stroop task, namely the task-relevant word and the task-irrelevant face.

Participants completed two functional localizer tasks, a 1-back task and a category judgment task, both of which were used to train classifiers. Two distinct localizer tasks were used to ensure that successful pattern classification was indeed contingent upon lower-level visual processing common across tasks and that the classifier was not driven by higher-level task demands. Trials were segregated into mini-blocks according to task-relevant information, creating 4 mini-block types: face, house, bird, and word. During face, house, and bird mini-blocks, participants were shown pictures of only faces, houses, birds, respectively, with non-words superimposed on either the top half or the bottom half of the image. During word mini-blocks, participants were shown a mixture of images of faces, houses, and birds, with real words superimposed on the image. Each mini-block lasted 8 s (4 TRs) and contained 8 trials, with each individual trial lasting 500 ms followed by a 500 ms inter-stimulus interval consisting of a fixation cross. Each block consisted of four mini-blocks, one for each stimulus type (i.e., faces, houses, birds, words) followed by an equal duration of fixation (i.e., 16 TRs/32 s). Four such blocks were presented for each task.

The logic of this design was to maximize the perceptual overlap between conditions of the localizer task (e.g. always a string of letters superimposed on an image) to allow the classifiers to identify activity corresponding to perceiving relevant information embedded with irrelevant information. This approach helps to de-correlate the classifiers, and it mirrors the situations on which the classifiers were tested (i.e., the Stroop tasks), thus increasing generalization performance (Lewis-Peacock and Norman, 2014).

During the 1-back localizer task, participants were instructed to press one of two buttons to indicate whether the task-relevant item on a given trial was the same or different from the task-relevant item on the previous trial. During the category judgment localizer task, participants were instructed to press one of two buttons to indicate whether the current task-relevant item belonged to one of two categories (Faces: Male/Female; Houses: With Deck/Without; Birds: Facing Left/Right; Words: Living/Non-living).

2.5.2.2. Stroop tasks. A spatial word – object Stroop task was used to identify regions involved in cognitive control as well as those involved in processing emotional information so the identified regions could be then used as ROIs for our multi-level modeling of the main task of interest, the emotional word – emotional face Stroop task (described

below). In the spatial-word – object Stroop task participants were instructed to make a two-choice manual response to the task-relevant word (e.g., “top”) based on its spatial meaning and to ignore its task-irrelevant location relative to the object on which it was superimposed. Participants pressed one button for words synonymous with “above” (“high”, “over”, “top”, “up”) and pressed another button for words synonymous with “below” (“low”, “under”, “bottom”, “down”). Words were positioned across the forehead (top positioned) on half the face trials and across the chin for the other half (bottom positioned). Words were positioned similarly on the houses, generally across the second story (top) on half the trials and across the front steps/porch on the other half (bottom). On congruent trials, the meaning of the word matched the word's location on the image (e.g. the word “over” on the top half of the image) while on incongruent trials the word's meaning and location did not match (e.g. the word “high” on the bottom half of the image). For half of the trials, the task-irrelevant picture was a house and for the other half, it was a face with a neutral expression. Trials with houses as the task-irrelevant dimension provided a baseline against which to evaluate increased activation when the task-irrelevant dimension was a neutral face. The regions so identified were then used as ROIs in our multi-level models to provide a proxy for brain regions involved in cognitive control and for those involved in emotional salience.

In the *emotional word – emotional face Stroop task*, individuals responded via a two-choice button press to indicate the emotional valence of the task-relevant word superimposed upon a task-irrelevant picture. Participants were instructed to press the right button for words synonymous with “sad” (“gloomy”, “miserable”, “sorrow”, “sad”) and the left button for words synonymous with “happy” (“cheerful”, “joy”, “delighted”, “happy”), while ignoring the emotional expression of the task-irrelevant face. On congruent trials the meaning of the word matched the emotional valence of the face (e.g. the word “miserable” on a sad face) and on incongruent trials the word's meaning did not match the emotional valence of the face (e.g. the word “joy” on a sad face) (refer back to Fig. 1). For half of the trials, the task-irrelevant item was a face with a happy expression and for the other half, a face with a sad expression. To maintain consistency with the spatial word – object Stroop task the location of the word varied, for half of the trials it was presented at the top (i.e., positioned across the forehead) and for the other half at the bottom (i.e., positioned across the chin).

A hybrid event related/blocked design was used, with two types of distractors; positively valenced faces and negatively valenced faces for the emotional word–emotional face Stroop task and neutral valenced faces and houses for the spatial word–object Stroop task. After a 23-second instruction screen, participants viewed 8 blocks of trials, each of which was subdivided into two halves that varied by the nature of the distractor (e.g., houses (A) versus neutral faces (B); sad faces (A) versus happy faces (B)) with an AB-BA-AB-AB-BA-BA-AB-BA order. Each half-block consisted of 10 trials. These half blocks consisted either predominantly of congruent trials or predominantly of incongruent trials. The remaining trials in each half block were composed of fixation trials and trials of the non-predominant condition. On average across each half-block, the predominant conditions accounted for 60% of the trials, the non-predominant condition accounted for 20% of the trials and fixations account for 20% of the trials. This design allowed for the potential of a blocked analysis, but here we focus instead on event-related analyses since we wish to examine the trial-by-trial covariation between brain activation and behavioral performance. Across all blocks, 40% of trials were congruent, 40% were incongruent, and 20% were fixation, with ordering optimized via optseq (<http://surfer.nmr.mgh.harvard.edu/optseq/>). Each block was preceded and followed by an 18.33 s block of fixation trials and a 4.67 s instruction reminder screen. Each trial lasted for 1950 ms(ms), followed by an inter-stimulus interval of 350 ms consisting of a fixation cross. Overall, there were 64 congruent trials, 64 incongruent trials, and 32 fixation trials per task.

2.5.3. Data pre-processing and analysis

2.5.3.1. Localizer tasks. For the 25 participants who had less than 3 mm of motions on both tasks, raw fMRI data volumes were individually motion corrected using FSL's MCLFIRT motion correction algorithm. For each subject, the resulting volumes for both tasks were concatenated together into one run, which was then motion corrected again using MCFLIRT to account for inter-run variations in head position. No smoothing was performed on the localizer data.

MVPA was performed within an anatomically-defined ROI (using the Harvard-Oxford Cortical Structural Atlas) selected to encompass major regions of the ventral visual processing stream, including early visual processing regions. This mask consisted of the following regions, all taken from the Harvard-Oxford Cortical Atlas and thresholded at a 20% probability level: intracalcarine cortex, lingual gyrus, lateral occipital cortex (inferior), occipital fusiform gyrus, occipital pole, parahippocampal gyrus (anterior and posterior divisions), temporal fusiform cortex (anterior and posterior divisions), temporal occipital fusiform cortex and inferior temporal gyrus (posterior).

Classification analyses were run using a penalized logistic regression classification algorithm, using L_2 (ridge) regularization with a penalty parameter of 50. Distinct classifiers were created for each participant individually, and were applied only to that participant's data. In order to remove non-informative voxels from the MVPA analyses, a feature-selection analysis of variance (ANOVA) was performed on each subject's preprocessed data, selecting only voxels whose activity varied significantly ($p < .05$) between stimulus image conditions (i.e., significant F for the comparison across birds, houses, words, and faces). Data from 32 s blocks of trials were used to train a classifier for each of four distinct visual categories: faces, houses, birds, words. The classifier used here was not multinomial (which could distinguish faces vs. houses vs. birds vs. words in a single classifier), but rather it was a binomial classifier that was used to construct four two-way classifiers, one for each category vs. the "other" three categories (faces vs. others, houses vs. others, birds vs. others, and words vs. others). The output from these classifiers were treated as category-specific estimates of the pattern activation for each category. As each block consisted of four 8-Section/4-TR long condition-specific mini-blocks (one mini-block per condition), all regressors were shifted forward in time by 4 s to account for the lag in the hemodynamic response as indexed by the BOLD signal.

To evaluate classifier training accuracy, we performed k-fold cross validation, in which k-1 32-second blocks (each of which contained a mini-block for each visual category) were used for training and the kth block used for testing, repeated for a total of k iterations, sampling without replacement, with each block being used for testing data once ($k = 8$). For every 2-sec TR of fMRI BOLD data, the classifier produced a "decoding" estimate (0–1) of the degree to which brain activity matched the category-specific pattern of brain activity learned by the classifier during training. For each TR, we determined which of the four classifiers provided the highest estimate (i.e., match) with brain activity

on that specific TR. This TR was then assigned to that visual category.

2.5.3.2. Stroop tasks

2.5.3.2.1. Standard GLM analyses. Univariate fMRI analyses were performed within FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>) for each Stroop task including pre-processing involving motion correction, brain extraction, high-pass filtering spatial normalization, and smoothing. For both Stroop tasks, the first 10 functional volumes (23 s) were discarded to allow the scanner to reach steady-state stability, after which MCFLIRT motion correction, BET brain extraction, and high-pass filtering (100 s) were applied. The resulting images were then registered and spatially normalized to the Montreal Neurological Institute (MNI) 152-T₁ 2-mm template. Individual subject's functional images were registered to their structural MPAGE images using 6-degree of freedom linear registration, which were in turn registered to MNI-space using 12-degree of freedom nonlinear registration. The resulting MNI-space images were smoothed with an 8-mm full-width at half-maximum Gaussian kernel.

FEAT was used to perform lower-level functional imaging statistics on each Stroop task separately. Using the general linear model (GLM), statistical maps were produced for each participant individually, showing associations between BOLD time-series data and linear regressors. Event-related regressors for accurate trials included congruent, incongruent, and fixation trials for both Stroop tasks, as well as neutral face and house distractor trials for the spatial word-object Stroop task, and sad face distractor and happy face distractor trials for the emotional word – emotional face Stroop task. Error trials were additionally included as a confound regressor. Because the mean RT for participants was substantially less than our 2.3 s TR, we modeled only the BOLD signal occurring between the onset of a trial and the button response on that trial. To do so we included RT regressors for all correct trials in seconds (i.e., variable epoch design). By modeling only the period of time prior to a response, we excluded brain activation associated with the period of time following a response but before the onset of the next trial, time during which participants may momentarily stop attending to the specific task demands. Head motion confound regressors of no interest were used for spikes in head motion greater than 3 mm within one volume, with the participant's head then returning to within 3 mm of the original starting position.

Higher-level group mixed effects models were then run using FLAME 1. One higher-level model was run without individual difference covariate regressors to investigate overall group effects. Then two separate models with a single individual differences covariate were run. The covariates employed were Z-transformed regressors representing each participant's EC and NA composites, respectively. All higher-level analyses were corrected for multiple comparisons via permutation testing using FSL's *randomise* function. For each contrast of interest, one-sample *t*-tests were performed using Threshold-Free Cluster Enhancement (TFCE). 5000 permutations were run for each contrast,

Table 1

Mean RT and accuracy for Stroop tasks by condition and image type. Mean and standard deviation (in parentheses) of RT and accuracy across subjects. Trials are divided up according their condition and distractor image type for both the spatial word-object Stroop task and emotional word-face Stroop task.

Stroop task	Condition	Image type	Mean RT (ms)	Mean Acc (%)
Spatial word – object Stroop task	Congruent	face	710.09 (86.88)	96.75 (6.05)
		house	711.17 (86.30)	96.75 (5.45)
		average	710.63 (84.97)	96.75 (5.48)
	Incongruent	face	725.25 (93.98)	94.75 (6.75)
		house	736.77 (79.62)	93.32 (5.48)
		average	731.01 (84.47)	94.04 (5.39)
Emotional word-face Stroop task	Congruent	negative	723.24 (95.30)	96.61 (4.65)
		positive	738.75 (104.10)	96.61 (4.30)
		average	731.07 (98.02)	96.51 (4.06)
	Incongruent	negative	753.49 (97.66)	94.86 (6.31)
		positive	762.41 (104.52)	93.64 (4.76)
		average	757.96 (99.15)	94.14 (5.12)

producing corrected 1-*p* maps, which served as the basis for all figures and Table 1-*p* was set at .95. Then, in additional familywise error correction ($p < .05$) was applied.

2.5.3.2.2. Multi-level modeling. We used multi-level modeling to examine the effect of brain activation on behavioral performance (RT) on a trial-by-trial basis. Multilevel mediation and moderation analyses were carried out using the Multilevel Mediation and Moderation (M3) toolbox for Matlab (Wager et al., 2009; http://wagerlab.colorado.edu/wiki/doku.php/help/mediation/m3_mediation_fmri_toolbox). By employing a multilevel approach, we were able to account for both first-level (within-subjects) and second-level (between-subjects) effects in a single structural model. In addition, we could examine the relationship between our variables of interest as they influence RT on accurate trials only. These models allowed us to test whether associations between the time series for a given variable (e.g., signal change in DLPFC) (X) and RT (Y) are mediated by other variables (i.e., classifier estimates of task-irrelevant face processing, classifier estimates of task-relevant word processing) (M). We used the FSL function *fsmeans* to extract time-series data from the emotional word-emotional face Stroop task for the ROIs of interest for each subject. Single-trial models were run separately for incongruent and congruent trials.

We tested a model with the following structure typical of mediation/moderation analyses. The *a* path (X – M) in the mediation models tested for relationships between the time-series of the percentage signal change of an ROI of interest (DLPFC, amygdala) (i.e., X) and the classifier fit for words and faces respectively (M₁ and M₂). Models for each ROI were run separately. The *b* path (M – Y) tested the relationships between classifier fits (M) and reaction time (Y), controlling for the effects of percentage signal change in our ROIs (X). The total effect, or *c* path (X – Y) tested for relationships between the time-series of percentage signal change for our ROIs (X) and RT (Y), without accounting for the effects of the classifier fits (mediation variables, M). The direct effect, or *c'* path, tested for relationships between the time-series of percentage signal change for our ROIs (X) and RT (Y), after accounting for the effects of the classifier fits for words and faces (M). Finally, the *ab* path (X – M – Y) tested for the mediating effect of the classifier fits (M) on the relationship between percentage signal change for our ROI (X) and RT (Y). To assess our assumption that DLPFC and amygdala are modulating perceptual processing, we also tested models with the X and M variables reversed (i.e., perceptual processing, as assessed by the classifier fit (X), influences RT (Y)), with DLPFC and amygdala activation acting as mediators. Consistent with our assumption, these models provided a poorer fit for the data.

All multi-level models were run on congruent and incongruent trials separately because task-irrelevant information has differential effects on RT for incongruent versus congruent trials. On incongruent trials, longer RTs likely indicate a reduced ability to ignore the task-irrelevant dimension as it conflicts with the response that should be emitted on the basis of the task-relevant dimension. The opposite is true on congruent trials – shorter RTs likely indicate a reduced ability to ignore the task-irrelevant dimension (MacLeod and MacDonald, 2000). In sum, four models were run: two for incongruent trials and two for congruent trials, and within each of these one using the DLPFC as our ROI of interest and another using the amygdala. In all models, the classifier fits for face and word processing were considered simultaneously.

For each of the four sets of models, three distinct models were run, one without any second-level moderators, one with EC as a second-level moderator, and one with NA as a second-level moderator. Analyses with second-level moderators investigated the degree to which the paths discussed above are altered by these variables relating to individual differences. We treated subject as a random factor, allowing both the slope and intercept to vary between subjects. To accurately characterize the distribution of the data, bootstrapping was carried out, using 10,000 bootstrap samples (Efron, 1979). In all models, path significance was determined in accordance with bias-corrected alpha

levels calculated through the bootstrapping procedure.

3. Results

3.1. Behavior

3.1.1. Stroop tasks

The raw RT and error data for the Stroop task is presented in Table 1. To control for individual differences in RT, we examined the percentage increase in RT on incongruent compared to congruent trials (RT for incongruent trials – RT for congruent trials/ RT for congruent trials) as our behavioral measure of interference. Significant interference was observed in the spatial word-object Stroop task (house distractors: mean difference = 0.039, $t(27) = 3.52$, $p < .001$, one-tailed, Cohen's $d = 0.66$; neutral face distractors: mean difference = 0.022, $t(27) = 2.18$, $p < .025$, one-tailed, Cohen's $d = 0.41$). The size of these effects did not vary by distractor type (houses vs. faces) ($t(27) = 1.37$, $p > .15$, two-tailed). Similarly, significant interference was observed in the emotional word – emotional face task (positive face distractors, mean difference = 0.034, $t(27) = 2.81$, $p < .01$, one-tailed, Cohen's $d = 0.53$; negative face distractors: mean difference = 0.044, $t(27) = 3.88$, $p < .0005$, one-tailed), Cohen's $d = 0.73$. The size of these effects did not vary by valence (positive vs. negative) ($t(27) = 0.618$, $p > .5$, two-tailed).

Accuracy in the spatial word – object Stroop task was significantly poorer on incongruent than congruent trials for house distractors (mean difference = -0.034, $t(27) = -4.35$, $p < .001$; one-tailed, Cohen's $d = 0.82$), but only approached significance for neutral face distractors (mean difference = -0.02, $t(27) = -1.85$, $p < .10$, Cohen's $d = 0.35$). However, the size of these effects did not significantly differ between distractors (house, face) ($t(27) = -0.952$, two-tailed). Accuracy in the emotional word – emotional face Stroop task was poorer on incongruent than congruent trials for positive face distractors (mean difference = -0.030, $t(27) = -3.13$, $p < .01$, one-tailed, Cohen's $d = 0.59$), but only approached significance for negative face distractors (mean difference = -0.018, $t(27) = -1.71$, $p < .10$, one-tailed, Cohen's $d = 0.32$). However, the size of these effects did not significantly differ between distractors (positive, negative) ($t(27) = -1.441$, $p > .15$, two-tailed). These results indicate that the task required cognitive control and that task performance was in line with expectations.

3.1.2. Questionnaires

The EC and NA composite scores were negatively correlated ($r = -0.502$, $df = 27$, $p < .01$, two-tailed) as has been reported previously (e.g., Snyder et al., 2015). Neither score correlated with the degree of interference for either Stroop task, nor with any demographic variables (IQ, age).

3.2. Neuroimaging

3.2.1. Localizer tasks

In order to evaluate whether classifiers trained on the localizer task data could accurately differentiate when a participant is attending to a face vs. a word, we first tested the classification accuracy of classifiers trained and tested solely on the localizer data. Given that we had four distinct categories, chance performance for the classifier was 25%. Classification performance for each category (birds, houses, words, faces) was significantly above chance (birds: mean = 71.8%, $t(24) = 14.00$, $p < .001$, one-tailed, Cohen's $d = 2.8$; houses: mean = 89.0%, $t(24) = 29.94$, $p < .001$, one-tailed, Cohen's $d = 6.0$; faces: mean = 73.6%, $t(24) = 15.29$, $p < .001$, one-tailed, Cohen's $d = 3.1$; words: mean = 48.8% $t(24) = 7.97$, $p < .001$, one-tailed, Cohen's $d = 1.6$). The confusion matrix for our classifiers is shown in Table 2. To help depict which voxels likely played a large role in discriminating between these faces and words, Fig. 2 depicts those voxels that yielded positive importance for one category (e.g., faces) and negative

Table 2

Percentage correct of MVPA classifications for each of the four different categories of visual stimuli used in the localizer scans. Each row represents the actual category of the stimuli and each column represents the percentage of those stimuli that were classified into each of the four visual categories. Values across rows equal 100%. Classifiers correctly identified each of the visual categories significantly above chance (i.e., 25%) as depicted in the diagonal of the table and highlighted in bold. Classification accuracies were determined by a leave one out procedure (out of 8 blocks total).

Classifier prediction				
	<i>face</i>	<i>house</i>	<i>bird</i>	<i>word</i>
Actual trial type				
<i>face</i>	73.6	4.6	11.4	10.4
<i>house</i>	1.2	89.0	3.8	6.0
<i>bird</i>	15.0	6.0	71.8	7.2
<i>word</i>	18.2	17.6	15.4	48.8

importance for the other (e.g., words).

3.2.2. Spatial word – object stroop task

The main goal of these analyses was to identify those brain regions related to cognitive control and emotional processing that are engaged specifically when faces must be ignored (i.e., when they are task irrelevant) as compared to ignoring other complex stimuli. To do so we identified those regions that exhibited greater activation when faces were irrelevant as compared to when houses were irrelevant (averaged over incongruent and congruent trials). This contrast yielded activation in three relevant ROIs: the right middle frontal gyrus (BA46, $x = 44$,

$y = 18$, $z = 20$, max $Z = 4.97$, $N = 137$ voxels), which is involved in cognitive control, the left amygdala (BA34, $x = -18$, $y = -6$, $z = -18$; max $Z = 7.85$, $N = 70$ voxels) (see Fig. 3), and a large cluster in the right fusiform gyrus with a peak in the right fusiform face area (BA37, $x = 44$, $y = -46$, $z = -24$; max $Z = 10.30$, $N = 417$ voxels). Ten mm voxel spheres centered around the peaks of the amygdala and middle frontal gyri clusters were used for the multi-level analyses of the emotional word – emotional face Stroop task.

Additionally, a ten mm voxel sphere centered around the peak of the fusiform gyrus cluster was used to index activity of the fusiform face area (FFA). To confirm that this ROI in fact captures the fusiform face area, we input the peak coordinates into NeuroSynth (neurosynth.org), an online tool that performs automated fMRI meta-analyses, and the top eight associated terms were either related to face or fusiform face area.

Since prior studies have used a GLM contrast (e.g., faces > houses) to identify regions whose activity might be altered by cognitive control, we examined the degree to which activity in this FFA ROI correlated with the fit of the face classifier used in the present study. To do so, we performed a repeated-measures correlation analysis on a trial-by-trial basis between activation in the FFA ROI and the classifier fit for faces, accounting for the non-independence caused by having multiple trials per subject. These results indicated that FFA activation was not significantly correlated with the fit of the face classifier ($r = -0.007$, $p = .717$) suggesting that our MVPA approach can detect neural processes to which standard univariate techniques are largely insensitive.

3.2.3. Emotional word – emotional face stroop task

3.2.3.1. Standard GLM analyses. The goal of these analyses was to confirm that our task and methods were tapping into the desired

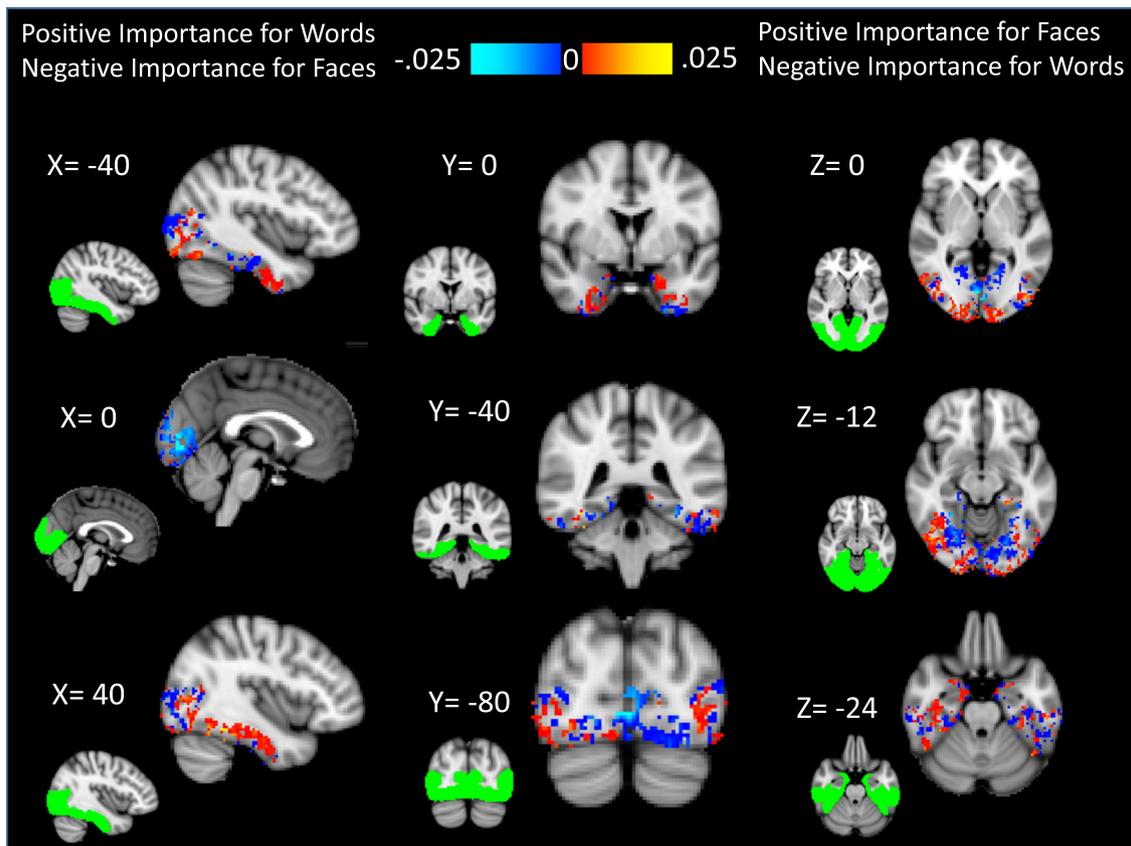


Fig. 2. Regions that show differential importance in contributing to the face and word classifiers. Shown here are the voxels with positive importance for one category classifier (e.g., faces), but with negative importance for the other (e.g., words). These voxels likely differentially contribute to the ability of the face and word classifiers to distinguish these two categories. Red indicates voxels that had positive importance for faces but negative importance for words, whereas blue indicates voxels that had positive importance for words but negative importance for faces. The spatial extent of the mask used for determining the classifier is shown in green.

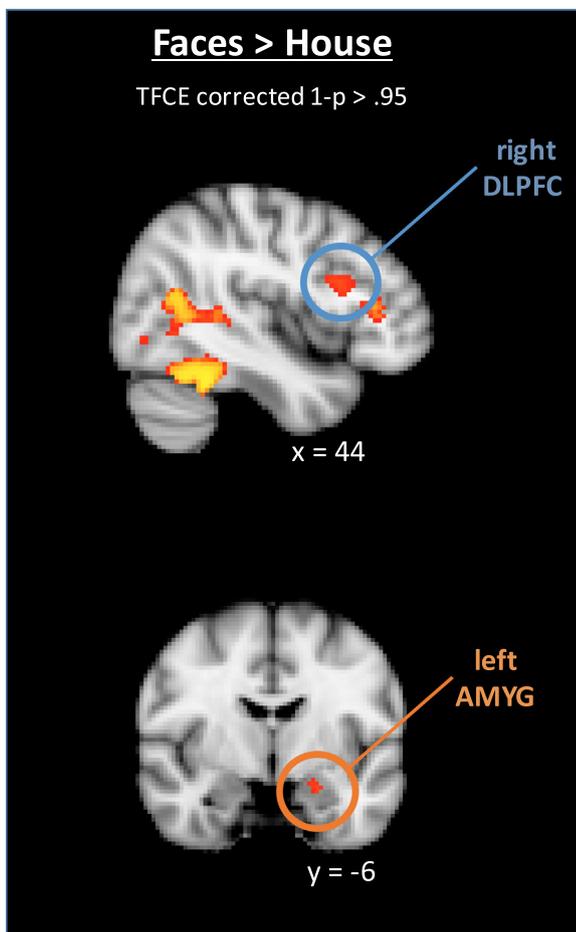


Fig. 3. Regions of the brain that served as ROIs for the multi-level modeling. These regions showed greater activation for trials with neutral face distractors as compared to than house distractors in the spatial word – object Stroop task at a TFCE corrected threshold of $1-p > .95$. (DLPFC= dorsolateral prefrontal cortex; AMYG = amygdala).

Table 3

Regions of significant activation in the emotional word-face Stroop task for the group as a whole as determined by a GLM analysis. All clusters were identified with Threshold Free Cluster Enhancement at $1-p$, with p at .95 and family wise error. BA = Brodmann area, Max Z = maximum Z score within the cluster, # of vox = number of voxels within the cluster x, y, z = MNI coordinates of cluster peak.

Region	BA	Max Z	# of vox	x	y	z
A. Incongruent & Congruent						
Incongruent > Fixation						
Inferior Occipital Gyrus(R)	BA17	15.3	6248	22	-94	-6
Inferior Occipital Gyrus(L)	BA18	16.3	5056	-28	-90	-6
Inferior Frontal Gyrus(R)	BA45	6.54	326	46	26	12
Insula(R)	BA13	5.74	105	44	20	-2
Precentral Gyrus(L)	BA6	5.45	73	-40	6	28
Parahippocampal Gyrus(R)	BA27	9.42	50	20	-30	-2
Parahippocampal Gyrus(L)	BA27	7.8	22	-20	-30	-6
Middle Frontal Gyrus(R)	BA6	4.16	14	50	8	38
Congruent > Fixation						
Fusiform Gyrus(R)	BA19	16	9926	36	-84	-12
Middle Frontal Gyrus(R)	BA 9	5.22	271	40	22	22
Parahippocampal Gyrus(R)	BA27	6.93	23	24	-28	-2
Caudate		5.86	15	20	10	0

constructs. As such, the main group effects are discussed here only briefly. Examination of activation for each trial type (incongruent, congruent) versus fixation baseline revealed a highly similar pattern of

activation across frontal cognitive control regions (extensive lateral PFC activation from inferior frontal regions back past the inferior frontal junction (IFJ)), operculum, bilateral superior parietal lobe, bilateral amygdala, as well as activation of posterior regions involved in processing faces (See Table 3). There were no significant differences in activation between the incongruent and congruent conditions, which, as we consider in the discussion, most likely results from the saliency to adolescents, in both the incongruent and congruent trials, of the task-irrelevant faces. With regards to differences driven by the valence of the face (positive vs. negative), no regions passed cluster correction with a whole-brain mask. But with a mask of the Faces > Houses contrast described above, significantly greater activation for negative than positive faces was observed in the right amygdala and in the right middle frontal gyrus, consistent with the idea that negatively valenced faces are more emotionally salient and require more cognitive control to ignore when such faces are irrelevant compared to positively valenced faces. This middle frontal gyrus region overlapped with the region identified in Incongruent > Fixation and Congruent > Fixation contrasts. We found no evidence that these patterns of activation were significantly influenced by either self-reported EC or NA.

3.2.3.2. Multi-level modeling. The goal of multi-level models was to examine how much the degree of activation in our two ROIs, one involved in cognitive control (right DLPFC) and the other involved in emotional processing (left amygdala), influenced RT on a trial-by-trial basis. The results for each of the three iterations (without any moderators, with EC as a moderator, with NA as a moderator) for each of our four models (Congruent/Incongruent trials by right DLPFC/left amygdala ROIs) are shown in Fig. 4 and Tables 4–7. It should be noted that the values for the pathways in the models without any moderators were essentially identical to those found in the models including moderators, hence only the former are reported.

3.2.3.2.1. Congruent trials. We first examined the relationship between activity in each of the relevant ROIs (DLPFC, amygdala) and RT. In neither the DLPFC model nor the amygdala model, did the level of activity in the relevant ROI (i.e., DLPFC, amygdala) predict RT, either when considered as a total effect (c path) nor as a direct effect with mediators in the model (c' path) on congruent trials (left hand column, Fig. 4). Nonetheless, increased activity in the DLPFC ROI was significantly associated with a decreased classifier fit for task-irrelevant faces ($Z = -3.37, p < .001, \alpha = 0.006$). While the direction of the association was similar in the amygdala model, the effect did not pass the bootstrap threshold ($Z = -2.33, p < .025, \alpha = 0.014$).

Then we examined the relationship between the degree of fit for each of the classifiers (i.e., word, face) and RT. The classifier fit from the task-relevant word in both the DLPFC and amygdala models were shown to influence RT ($b1$ paths) after taking DLPFC ($Z = 3.11, p < .0025, \alpha = 0.006$) and amygdala ($Z = 2.65, p < .01, \alpha = 0.014$) activation into account, respectively. The direction of this effect was that increased word processing was associated with slower RTs. For neither model was there a significant effect of face processing on RT ($b2$ path). The lack of this significant $b2$ path may have resulted from opposing moderating effects of individual differences, which we discuss next.

We also examined the degree to which the pathways were influenced by individual differences in EC and NA. In particular, higher person-level EC scores were significantly associated with a decreased influence of the face classifier fit on RT in both the DLPFC ($Z = -3.41, p < .0007, \alpha = 0.006$) and amygdala ($Z = -3.39, p = .0007, \alpha = 0.0146$) models (See Fig. 5). In contrast, increased person-level NA was significantly associated with an increased influence of the face classifier fit on RT in the amygdala model ($Z = 2.64, p < .009, \alpha = 0.011$) and in the same direction in the DLPFC model, but did not pass the bootstrap threshold ($Z = 2.49, p < .013, \alpha = 0.006$).

None of the ab (i.e., marginal mediation; indirect effects) pathways

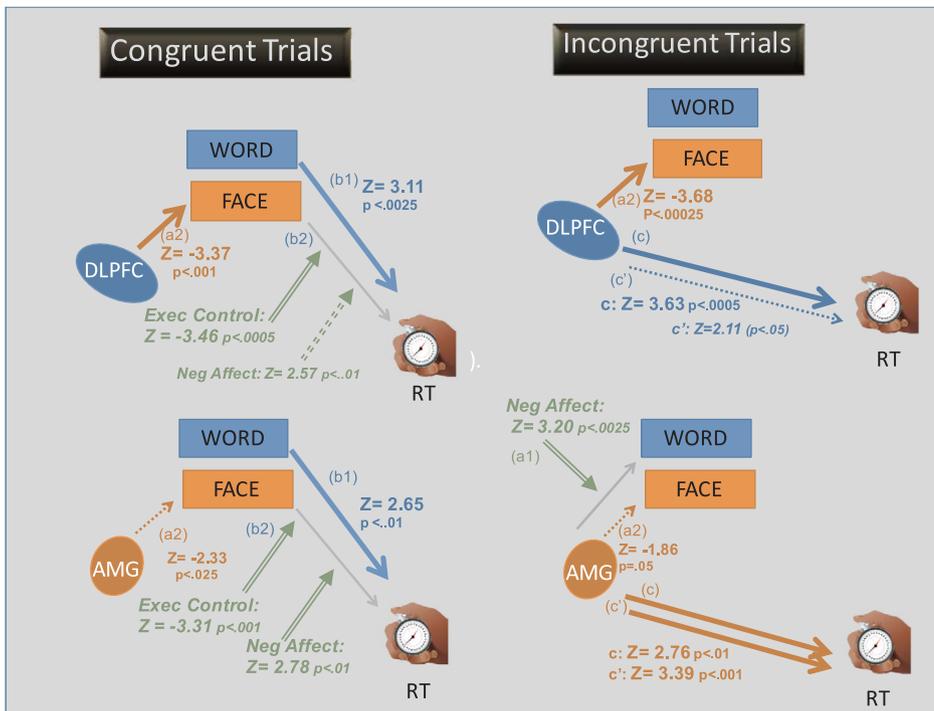


Fig. 4. The outcome of the mediation/moderation models predicting trial-by-trial performance. Top Row: The relationships between DLPFC activity and reaction time. Bottom Row: The relationships between amygdala (AMYG) activity and reaction time. Left hand column: The relationships for congruent trials. Right hand column: The relationships for incongruent trials. Trial-by-trial Fits for word and face classifiers, respectively serve as potential mediators in the model. Significant effects that pass the bootstrapping threshold are shown by solid arrows. Those that do not pass the bootstrapping threshold are shown by dotted arrows. Brain regions involved in cognitive control and processing task-relevant information are shown in blue. Brain regions involved in emotional salience and processing task-irrelevant information are shown in orange. The effects of individual differences moderators are shown in green and are depicted by arrows with a double line shaft. Notable findings are that 1) increased DLPFC activation in both incongruent and congruent trials was associated with a reduce fit for the face classifier, consistent with top-down modulation to decrease the influence of the task-irrelevant face, 2) both DLPFC and AMYG activation influence RT on incongruent but not congruent trials, with the direct effect only significant for the AMYG model,

and 3) that individual differences moderate the effect of how much processing of the task-irrelevant face influences RT, being increased for individuals with high negative affect and decreased for individuals with high executive control. (DLPFC = dorsolateral prefrontal cortex; AMYG = amygdala; RT = reaction time).

reached significance as determined by the bootstrapped alpha level. Nonetheless, EC also moderated the indirect path between DLPFC activation, word classifier fit, and RT (*a1b1* path) ($Z = -4.25, p < .0001, \alpha = 0.006$). However, this effect is difficult to interpret given the lack of significance of the *a1* and *b1* pathways in this model (See Tables 4, 5).

3.2.3.2.2. Incongruent trials. With regards to the effect of activity in each of the ROIs of interest (right hand column, Fig. 4), there were significant total effects of DLPFC and amygdala activity on RT (*c* paths), such that increased activity was associated with elongated RT (DLPFC model: $Z = 3.63, p < .001, \alpha = 0.006$; amygdala model: $Z = 2.76, p < .006, \alpha = 0.006$). After taking into account the other mediators in

Table 4

Moderated mediation results for the right DLPFC models across congruent trials. Three sets of models were run: one without moderators, one with EC as a moderator and one with NA as a moderator. The results for each model are presented separately within the section for each outcome variable. Significant paths, which are those whose *p* value exceeds the bias-corrected alpha value as determined by bootstrapping, are shown in bold.

Moderated mediation results for congruent trials – DLPFC models							
Outcome variable	Predictor	path	<i>b</i>	SE	<i>Z</i>	<i>p</i>	α
word	dlpfc	a1	2.03	1.25	1.67	.0956	0.006
	dlpfc × EC	a1 × EC	− 0.59	1.48	− 0.78	.4325	0.006
	dlpfc × NA	a1 × NA	0.98	1.80	0.51	.6129	0.006
face	dlpfc	a2	− 3.12	0.9	− 3.37	.0007	0.006
	dlpfc × EC	a2 × EC	0.33	1.12	0.26	.7957	0.006
	dlpfc × NA	a2 × NA	− 0.28	1.38	− 0.60	.5517	0.006
RT	dlpfc	<i>c</i>	1	0.56	1.67	.0946	0.006
	dlpfc	<i>c'</i>	0.64	0.63	1.03	.3015	0.006
	dlpfc × EC	<i>c</i> × EC	− 0.85	0.63	− 1.48	.1378	0.006
	dlpfc × NA	<i>c</i> × NA	− 0.11	0.64	− 0.23	.8215	0.006
	dlpfc × EC	<i>c'</i> × EC	− 0.50	0.68	− 0.43	.6700	0.006
	dlpfc × NA	<i>c'</i> × NA	− 0.57	0.71	− 0.91	.3623	0.006
	word	b1	0.08	0.03	3.11	.0019	0.006
	word × EC	b1 × EC	− 0.02	0.04	− 0.67	.5013	0.006
	word × NA	b1 × NA	0.04	0.06	0.73	.4177	0.006
	face	b2	0	0.03	− 0.04	.9650	0.006
	face × EC	b2 × EC	− 0.12	0.03	− 3.41	.0006	0.006
	face × NA	b2 × NA	0.11	0.04	2.49	.0128	0.006
	Indirect effects						
dlpfc – word – RT		ab1	0.1	0.07	1.44	.1492	0.006
dlpfc – face – RT		ab2	0	0.07	− 0.03	.9780	0.006
dlpfc – word – RT × EC		ab1 × EC	− 0.26	0.13	− 4.25	< .0001	0.006
dlpfc – word – RT × NA		ab1 × NA	0.19	0.17	1.09	.1301	0.006
dlpfc – face – RT × EC		ab2 × EC	0.07	0.11	0.72	.4685	0.006
dlpfc – face – RT × NA		ab2 × NA	− 0.07	0.12	− 0.34	.7346	0.006

Table 5

Moderated mediation results for the left amygdala models across congruent trials. Three sets of models were run: one without moderators, one with EC as a moderator and one with NA as a moderator. The results for each model are presented separately within the section for each outcome variable. Significant paths, which are those whose p value exceeds the bias-corrected alpha value as determined by boot-strapping, are shown in bold.

Moderated mediation results for congruent trials – amygdala models							
Outcome variable	Predictor	path	b	SE	Z	p	α
word	amyg	a1	0.3	1.03	0.3	.7617	0.0137
	amyg × EC	a1 × EC	0.23	1.22	0.28	.7764	0.0146
	amyg × NA	a1 × NA	0.59	1.76	0.33	.7437	0.011
face	amyg	a2	− 1.43	0.59	− 2.33	.0197	0.0137
	amyg × EC	a2 × EC	0.30	0.80	0.22	.8295	0.0146
	amyg × NA	a2 × NA	− 0.65	0.90	− 1.10	.2729	0.011
RT	amyg	c	0.36	0.28	1.50	.1340	0.0137
	amyg	c'	0.06	0.35	0.16	.8724	0.0137
	amyg × EC	c × EC	0	0.36	0.29	.8295	0.0146
	amyg × NA	c × NA	− 0.10	0.52	− 0.01	.9883	0.011
	amyg × EC	c' × EC	0.28	0.42	0.68	.4702	0.0146
	amyg × NA	c' × NA	− 0.55	0.69	− 0.53	.5951	0.011
	word	b1	0.08	0.04	2.65	.0080	0.0137
	word × EC	b1 × EC	− 0.05	0.05	− 1.10	.2703	0.0146
	word × NA	b1 × NA	0.07	0.07	1.14	0.2536	0.011
	face	b2	− 0.01	0.03	− 0.28	.7786	0.0137
	face × EC	b2 × EC	− 0.13	0.04	− 3.39	.0007	0.0146
	face × NA	b2 × NA	0.12	0.05	2.64	.0083	0.011
	Indirect effects						
amyg – word – RT		ab1	0.02	0.05	0.51	.6132	0.0137
amyg – face – RT		ab2	− 0.03	0.05	− 0.49	.6268	0.0137
amyg – word – RT × EC		ab1 × EC	− 0.10	0.11	− 0.70	.4851	0.0146
amyg – word – RT × NA		ab1 × NA	0.13	0.12	1.19	.2330	0.011
amyg – face – RT × EC		ab2 × EC	− 0.05	0.06	− 1.38	.1665	0.0146
amyg – face – RT × NA		ab2 × NA	0.02	0.07	0.27	.7898	0.011

the model, however, only the amygdala showed a significant direct effect on RT ($Z = 3.39, p < .001, \alpha = 0.006$), whereas the direct effect of DLPFC activity on RT did not pass the bootstrap threshold ($Z = 2.11, p < .035, \alpha = 0.006$) (c' paths).

With regards to the effect of ROI activity on classifier fits, while there was no overall significant effect of amygdala or DLPFC activation

on the word classifier fit ($a1$ paths), the influence of this pathway moderated by an individual's level of NA ($Z = 3.05, p < .003, \alpha = 0.006$) in the amygdala model, indicating that for individuals with higher levels of NA, increased amygdala activity was associated with a higher fit for the word classifier.

However, there was a significant effect of DLPFC activity on the fit

Table 6

Moderated mediation results for the right DLPFC models across incongruent trials. Three sets of models were run: one without moderators, one with EC as a moderator and one with NA as a moderator. The results for each model are presented separately within the section for each outcome variable. Significant paths, which are those whose p value exceeds the bias-corrected alpha value as determined by boot-strapping, are shown in bold.

Moderated mediation results for incongruent trials – DLPFC models							
Outcome variable	Predictor	path	b	SE	Z	p	α
word	dlpfc	a1	4.31	2.97	1.37	.1721	0.006
	dlpfc × EC	a1 × EC	− 2.22	3.17	− 1.06	.2909	0.006
	dlpfc × NA	a1 × NA	− 1.99	4.89	− 0.09	.9319	0.006
face	dlpfc	a2	− 5.02	1.10	− 3.68	.0002	0.006
	dlpfc × EC	a2 × EC	1.56	1.59	0.97	.3306	0.006
	dlpfc × NA	a2 × NA	− 0.82	2.02	− 0.75	.4536	0.006
RT	dlpfc	c	1.92	0.063	3.63	.0003	0.006
	dlpfc	c'	1.25	0.66	2.11	.0346	0.006
	dlpfc × EC	c × EC	1.20	1.07	1.01	.3101	0.006
	dlpfc × NA	c × NA	− 0.42	1.30	− 0.76	.4457	0.006
	dlpfc × EC	c' × EC	1.28	1.12	1.02	.3061	0.006
	dlpfc × NA	c' × NA	− 0.40	1.42	− 0.67	.5054	0.006
	word	b1	0.02	0.02	1.59	.1121	0.006
	word × EC	b1 × EC	0.02	0.02	1.31	.1906	0.006
	word × NA	b1 × NA	− 0.04	0.02	− 2.37	.0176	0.006
	face	b2	− 0.04	0.02	− 2.13	.0328	0.006
	face × EC	b2 × EC	0.03	0.02	1.42	.1551	0.006
	face × NA	b2 × NA	− 0.05	0.03	− 1.86	.0634	0.006
	Indirect effects						
dlpfc – word – RT		ab1	0.13	0.1	1.35	.1786	0.006
dlpfc – face – RT		ab2	0.04	0.07	0.56	.5739	0.006
dlpfc – word – RT × EC		ab1 × EC	− 0.05	0.16	− 0.63	.5255	0.006
dlpfc – word – RT × NA		ab1 × NA	− 0.01	0.2	− 0.3	.7627	0.006
dlpfc – face – RT × EC		ab2 × EC	− 0.1	0.11	− 0.65	.5183	0.006
dlpfc – face – RT × NA		ab2 × NA	0.13	0.14	1.40	.1628	0.006

Table 7

Moderated mediation results for the left amygdala models across incongruent trials. Three sets of models were run: one without moderators, one with EC as a moderator and one with NA as a moderator. The results for each model are presented separately within the section for each outcome variable. Significant paths, which are those whose p value exceeds the bias-corrected alpha value as determined by boot-straping, are shown in bold.

Moderated mediation results for incongruent trials – amygdala models							
Outcome variable	Predictor	path	b	SE	Z	p	α
word	amyg	a1	- 0.39	2.38	- 0.2	.8429	0.006
	amyg × EC	a1 × EC	0.56	2.73	0.30	.7612	0.006
	amyg × NA	a1 × NA	7.20	3.45	3.05	.0023	0.006
face	amyg	a2	- 2.01	1.06	- 1.86	.0631	0.006
	amyg × EC	a2 × EC	0.76	1.30	0.41	.6805	0.006
	amyg × NA	a2 × NA	- 1.91	1.75	- 1.07	.2840	0.006
RT	amyg	c	1.22	0.47	2.76	.0057	0.006
	amyg	c'	1.28	0.43	3.39	.0007	0.006
	amyg × EC	c × EC	0.95	0.52	1.67	.0942	0.006
	amyg × NA	c × NA	- 0.34	0.74	- 0.27	.7869	0.006
	amyg × EC	c' × EC	0.85	0.50	1.59	.1125	0.006
	amyg × NA	c' × NA	- 0.74	0.63	- 1.06	.2884	0.006
	word	b1	0.03	0.02	1.69	.0911	0.006
	word × EC	b1 × EC	0.02	0.02	1.25	.2127	0.006
	word × NA	b1 × NA	- 0.04	0.02	- 1.92	.0552	0.006
	face	b2	- 0.04	0.02	- 2.11	.0348	0.006
	face × EC	b2 × EC	0.02	0.02	1.09	.2763	0.006
	face × NA	b2 × NA	- 0.05	0.03	- 1.47	.1427	0.006
	Indirect effects						
amyg – word – RT		ab1	- 0.17	0.09	- 1.83	.0678	0.006
amyg – face – RT		ab2	0.11	0.06	2.50	.0123	0.006
amyg – word – RT × EC		ab1 × EC	0.05	0.07	0.91	.3645	0.006
amyg – word – RT × NA		ab1 × NA	0.11	0.14	1.11	.2662	0.006
amyg – face – RT × EC		ab2 × EC	- 0.03	0.07	- 0.31	.7531	0.006
amyg – face – RT × NA		ab2 × NA	0.02	0.07	0.26	.7869	0.006

of the face classifier ($Z = -3.68, p < .00025, \alpha = 0.006$) ($a2$ path) such that increased DLPFC activity was associated with a decreased fit for the face classifier, consistent with top-down cognitive control over task-irrelevant information. A similar effect was observed in the amygdala model, but it was non-significant after bootstrapping ($Z = -1.86, p = .06, \alpha = 0.006$). For all four models, none of the ab (i.e., marginal mediation; indirect effects) reached significance.

With regards to the influence of the classifier fits on RT, no

significant effects were observed. In addition, none of the ab (i.e., marginal mediation; indirect effects) pathways reached significance (See Tables 6, 7).

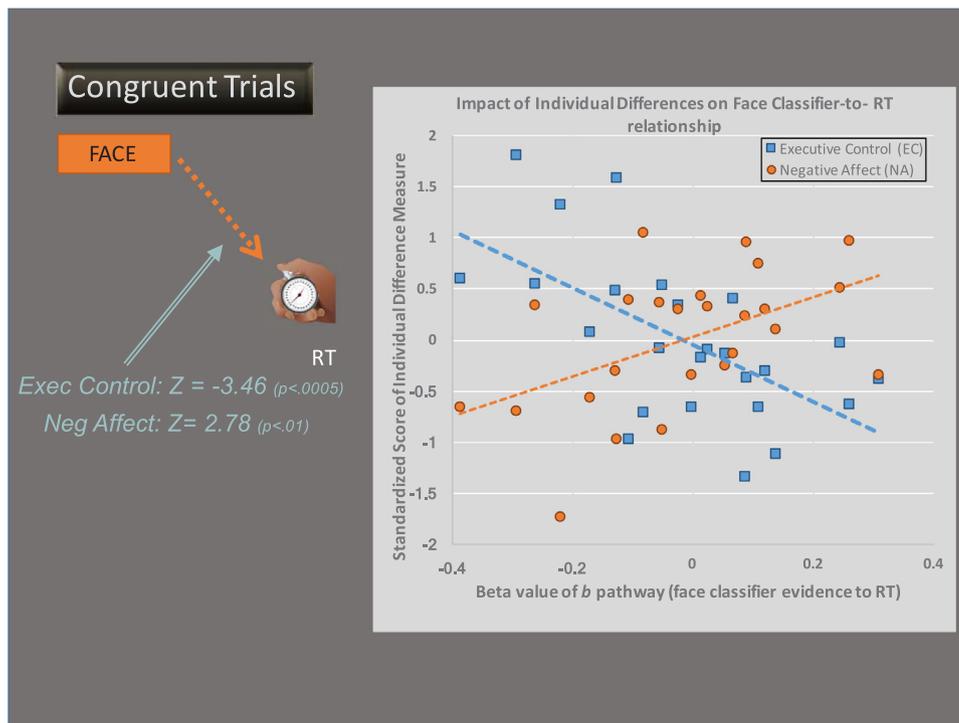


Fig. 5. Scatter plot of moderation effects of the b (face classifier evidence to RT) pathway. Data shown are for the model considering DLPFC effects on RT, although the results for the model considering amygdala effects on RT are practically identical. Levels of Executive Control are shown in blue and levels of Negative Affect are shown in orange. Higher levels of Executive Control are associated with a reduced effect of face classifier evidence on RT, while the opposite is true for higher levels of Negative Affect, as they are associated with an increased effect of face classifier evidence on RT.

4. Discussion

4.1. Overview

The current study used a multi-faceted analysis approach with fMRI data to investigate the degree to which cognitive control acts to modulate the processing of task-relevant as compared to task-irrelevant information in adolescents. Integrating standard univariate fMRI analysis techniques with sophisticated multivariate fMRI analyses and multi-level modeling, we found evidence that activity in a cognitive control region, the DLPFC, exerted its influence by modulating processing of the task-irrelevant facial information rather than by modulating task-relevant word processing, an effect observed for both incongruent and congruent trials. This finding is consistent with the findings on adult samples of Polk et al. (2008) and Coste et al. (2011), although in both those studies they also found modulation of processing of the task-relevant dimension as well. In our study, processing of the task-relevant information was associated with performance, but only for congruent trials and only directly (i.e., it not modulated by processing in other regions). At the same time, the direct influence of the task-irrelevant information on RT was moderated by individual differences in EC and NA. We now discuss these effects in more detail. Before doing so however, we discuss data supporting the underlying assumption of our paradigm, which is that the processing of faces during adolescence is particularly pre-potent and salient.

4.2. Prepotency of emotional faces

One of the underlying assumptions of our task design was that the task-irrelevant dimension of our stimuli, faces with emotional expression, would be pre-potent. This idea was based on findings that adolescence is a time period during which emotional information is particularly salient (e.g., Crone and Dahl, 2012), and leads to activation of brain regions involved in emotion processing (e.g., Hare et al., 2008; Guyer et al., 2008) even when such emotional information is task-irrelevant (Grose-Fifer et al., 2013; Monk et al., 2003). In fact, there was clear evidence of the prepotency of emotional information in our participants as deduced from the non-emotional spatial word-object Stroop task that we used to select our ROIs for the main analysis of interest. In this task, individuals made a decision about whether a word's meaning with regards to the concept of "above" or "below", while ignoring the spatial position of the word on the background of either a task-irrelevant house or a task-irrelevant face with a neutral expression. Compared to the processing of task-irrelevant houses, task-irrelevant neutral faces activated the medial amygdala, a region involved in processing the social/affective aspects of faces (Bickart et al., 2012). Importantly, amygdala activation facilitates attention to emotional information and is thought to act as a "first alert" system for the quick and efficient processing of emotional information (see Phelps, 2005 for a review), consistent with our reasoning that such information would be pre-potent to our participants.

4.3. Effects common to congruent and incongruent trials

Probably the most important finding of the current study (and also the strongest) is that increased activity in a control region, the DLPFC, was associated with a decreased classifier fit for the task-irrelevant facial information. This effect was observed across both incongruent and congruent trials, and suggests that the down-regulation of task-irrelevant information is a mechanism whereby control can be exerted. This finding is consistent with the larger body of research reviewed in the introduction suggesting that DLPFC modulates the processing of information with regards to its task relevance. Moreover, at least one study using a GLM approach with a task-irrelevant dimension similar to ours (emotional faces) has revealed that increases in cognitive control regions are associated with reduced activation in ventral visual areas

associated with face (fusiform face area) and object (extrastriate areas) processing (Steinhauser et al., 2016).

Some aspects of our results suggest that these effects are specific to control regions. We only found a marginal relationship between amygdala activation and the classifier fit faces based on a standard *p*-value and these relationships did not pass the more stringent requirements of permutation testing. Moreover, DLPFC and amygdala activation were positively correlated, and we speculate that the marginal effects for the amygdala are driven by its association with DLPFC activity. It would not be surprising to find such a positive association, as to the degree that there is increased amygdala activity on a given trial, DLPFC activity may need to be increased to modulate its influence.

4.4. Effects specific to congruent trials

In interpreting these results, it is important to remember that while poor cognitive control on a given trial leads to more processing of task-irrelevant information, regardless of trial type, the influence of control on RT differs by trial type. Poor control leads to increased RT on incongruent trials, but *decreased* RT on congruent trials. More specifically, if an individual is failing to exert cognitive control on a congruent trial, he or she will pay more attention to the task-irrelevant (but pre-potent) stimulus information. This will lead to a decrease in RT because the automatic processing of the task-irrelevant information aligns with and aids in the response to the task-relevant information. In contrast, if an individual is paying attention to the task-relevant dimension of a stimulus, RT will be elongated, as the dimension to which they are correctly directing attention, the task-relevant one, is not pre-potent and takes longer to process (see MacLeod and MacDonald, 2000, for a longer discussion).

Unique to congruent trials, we found that an increased classifier fit for the processing of the task-relevant words was associated with increased RT. As mentioned above, increased RT on a given congruent trial suggests good compliance with task demands by focusing on the task-relevant but less pre-potent dimension of a stimulus. However, and of importance, this effect was not influenced by trial-to-trial activity in the DLPFC, suggesting that the effect may not be driven by control processes. Consistent with this conjecture, the effect was observed in both the DLPFC and amygdala models. Obviously, however, one cannot draw conclusions from null results, and so this conclusion must remain speculative. It is also important to remember that our analysis was specifically examining trial-by-trial variation, and that congruent trials require less control than incongruent trials. As such, it may be that the overall task set employed across all congruent trials is adequate to keep the focus on the task-relevant word, and that variations in this top-down bias trial-to-trial make a minimal contribution to RT. Rather, it may be that RT is influenced by the degree of perceptual processing of that word, with trials for which there is difficulty in determining a response (as evidenced by longer RT) requiring a higher-level of perceptual processing (as evidenced by a higher classifier fit).

4.5. Effects specific to incongruent trials

The pattern that emerged for incongruent trials, which require higher levels of cognitive control than congruent trials, is somewhat different than observed for congruent trials. For incongruent trials, we observed that increased activation in DLPFC was significantly associated with increased RT (total effect). However, the direct effect was only marginal once taking into account other paths in the model. As such, it appears that the DLPFC-RT relations likely can be accounted for by the influence of DLPFC in downregulating processing of the task-irrelevant face (an effect observed for congruent trials as well).

Amygdala activation also had a direct effect on RT, but this remained when taking other pathways into account, suggesting that this effect is not mediated by processing of the task-relevant and task-irrelevant dimensions of the stimulus. As such, the contrast between the

pattern for the DLPFC and amygdala model provides evidence that the DLPFC is exerting its influence by modulation of the processing of the task-irrelevant information, while the amygdala is not.

4.6. Individual differences

Importantly, we found that differences amongst individuals in their self-reported Executive Control (EC) and Negative Affect (NA) moderated some of the effects we observed. While there was no significant overall effect on RT of the degree to which the task-irrelevant face is processed, individual differences moderated this relationship. As expected, EC and NA had opposing effects. Specifically, increasing levels of EC are associated with a *reduced* influence of processing of the task-irrelevant face (as indexed by classifier fit) on RT. Conversely, increasing levels of EC are associated with an *increasing* influence of the processing on RT. Both of these relationships are consistent with the argument made by MacLeod and MacDonald (2000) that poor control on congruent trials is associated with attention to the task-irrelevant dimension. We should note that our multi-level modeling was sensitive to these effects while the standard GLM analysis was not.

With regards to incongruent trials, only one significant effect was observed, which was specific to the amygdala model. In this model, an increased level of NA amongst individuals was associated with a stronger positive relationship between amygdala activity and the classifier fit for the word information. This finding suggests that the higher the level of NA, the more increased amygdala activity is associated with increased processing of the task-relevant word. We speculate that with higher levels of NA, attentional capture by emotional information is increased and extends not only to the prepotent emotional face, but also to the word because of its emotional nature. In support of this, previous research has found that subclinical levels of anxiety and depression, which are both highly associated with NA, are associated with the degree of amygdala responsiveness to emotional words (Laeger et al., 2012).

It is of interest that the influence of individual differences was observed mainly for congruent but not incongruent trials. We speculate that the effects of individual differences may have been more pronounced on congruent trials because there is more flexibility in the strategy used to emit a correct response on congruent trials. Making a correct response on incongruent trials requires focusing on the task-relevant dimension and reducing interference from the task-irrelevant dimension. As discussed by Kane and Engle (2003), when incongruent trials occur with high frequency, as they did within our half-blocks, the act of making a correct decision on incongruent trials, in and of itself, can serve as an intrinsic and continual reminder of the task goal: pay attention to the task-relevant dimension and not the task-irrelevant one. In contrast, no such intrinsic reminding is provided by congruent trials, as individuals can produce the correct decision either by focusing on the task-relevant dimension or by focusing on the task-irrelevant dimension. As such, congruent trials may be more sensitive to individual differences.

4.7. Limitations and future directions

As the first study to take such a multi-level approach to investigating the question under examination, there are, by necessity, a number of potential limitations. One potential limitation is that our distracting information was of an emotional nature, and as such may not generalize to non-emotional information. Whereas at least some control mechanisms appear to be shared in common across Stroop tasks with task-irrelevant emotional information as well as those with task-irrelevant non-emotional information, there are distinct mechanisms as well (e.g., Compton et al., 2003; Kaiser et al., 2015). As such, we might find that similar, distinct, or an expanded set of mechanisms (i.e. effects on both task-relevant information and task-irrelevant information) might be observed. One future issue that might be addressed, but may

be potentially difficult to disentangle, will be the degree to which these control mechanisms are influenced by the type of task-irrelevant material (emotional vs. non-emotional) as compared to the degree of control required. For example, the behavioral interference effect in the standard color-word Stroop task is usually quite robust (on the order of 70–100 ms) whereas effect on emotional Stroop tasks, such as those used in the present study are more modest (on the order of 30 ms).

Another potential limitation of the current study is that it investigates how cognitive control influences the processing of task-relevant and task-irrelevant information in a relatively narrow developmental time range, that of mid-adolescence (ages 14–17). It may be that we observed that control regions modulate the processing of the task-irrelevant information as to influence performance, precisely because such information, emotional faces, is extremely salient in the particular age range of our participants. As such, it is possible that during this developmental time period, the critical factor for obtaining good performance on our task is the degree to which control can be exerted to reduce the processing of task-irrelevant information, as compared to any gains in performance that might be obtained by upregulating processing of the task-relevant word. However, as such emotional facial information is likely not as salient for older (and younger) aged individuals, it may be that the more effective method for control regions to influencing performance in these age groups might be to modulate processing of the task-relevant dimension. In addition, as cognitive control regions are still developing during this period of adolescence (Andrews-Hanna et al., 2011), the pattern we observed in this age group might not reflect a mature or optimized pattern of control, which might involve expanded or distinct mechanisms. We currently are performing a follow-up study to address this issue in which we will have both adolescent and adult participants.

Nonetheless, we purposely focused on this age range because of the future potential for translation to issues relevant to psychopathology. During mid-adolescence, psychopathology related to depression and anxiety begin to manifest in earnest (e.g., Paus et al., 2008). Hence, it would be of interest to know whether the control mechanisms observed in our sample generalize to adolescents at substantially increased risk for psychopathology by virtue of family history (e.g., a mother with a history of depression) or life experiences (e.g., a trauma history). In such individuals, either the activation of cognitive control regions, emotional processing regions, or that of both might add additional independent predictive power in explaining the ability to ignore emotionally distracting information. This potential is hinted at by the fact that at least some portion of our results is indeed influenced by these individual differences dimensions. We are currently performing a follow-up study to examine this possibility.

Another limitation relates to the temporal dynamics of control. Though our characterization of brain-behavior relationships on a trial-by-trial basis may better capture the temporal dynamics of control than traditional GLM analyses, with a TR of 2000 ms, we are still limited in the fine-scale temporal dynamics of brain systems that are likely influential on trial-by-trial task performance. For instance, it may be that DLPFC activity at the beginning of a trial (e.g. the first 500 ms of a trial) may be associated with increased task-relevant stimulus processing, but not at the end of a trial. As such, a temporal resolution of 2000 ms, as used in the current Stroop paradigms, may obscure important brain-behavior relationships that are specific for different points during a given trial. Ongoing follow-up research in our laboratory is employing a TR of 460 ms, allowing for greater temporal resolution to evaluate this potential issue.

Finally, while the findings from the multi-level mediation-moderation analyses are quite intriguing, it must be noted that an N of 25 is relatively limited for an individual differences study. As such, it would be helpful to explore whether these effects replicate in a separate and larger sample, an endeavor in which we are currently engaged.

4.8. Conclusions

By taking a systematic, carefully controlled and multi-faceted analysis approach, the current study provided unique information on how cognitive control influences the processing of task-relevant and task-irrelevant information. Leveraging the power of MVPA, we showed that on a trial-by-trial basis, DLPFC acts to exert control by influencing processing of the task-irrelevant material, in this case a face with an emotional expression, in an emotional word – emotional face Stroop task. In contrast to prior studies, no effect on the processing of task-relevant information was found. Furthermore, by comparing our model for DLPFC influences to that of the amygdala, we were able to show moderate specificity of these effects. In addition, our study provides evidence that the influences on RT are affected by individual differences in cognitive control and affective processing, specifically EC and NA, respectively. Such findings have important implications for notable health-related issues that are relevant to individuals during mid-adolescence, including the onset of psychopathology. They suggest that well beyond measuring an individual's tendency towards negative affect, understanding the development of their cognitive control abilities may be quite beneficial, as these abilities may help to buffer against the emotional distractions and turmoil of youth.

Acknowledgements

The authors declare no competing financial interests. The work reported in this paper was enabled by a James McKeen Cattell sabbatical award to the first author and support from the National Institute of Mental Health R01MH105501 to M. Banich and B. Hankin (MPIs).

Declaration of interests

None.

References

- Adelman, N.E., Menon, V., Blasey, C.M., White, C.D., Warsofsky, I.S., Glover, G.H., Reiss, A.L., 2002. A developmental fMRI study of the stroop color-word task. *NeuroImage* 16 (1), 61–75. <https://doi.org/10.1006/nimg.2001.1046>.
- Aloi, J., Blair, K.S., Crumb, K.I., Meffert, H., White, S.F., Tyler, P.M., et al., 2018. Adolescents show differential dysfunctions related to alcohol and cannabis use disorder severity in emotion and executive attention neuro-circuitries. *NeuroImage-Clin.* 19, 782–792. <https://doi.org/10.1016/j.nicl.2018.06.005>.
- Andrews-Hanna, J.R., Mackiewicz Seghete, K.L., Claus, E.D., Burgess, G.C., Ruzic, L., Banich, M.T., 2011. Cognitive control in adolescence: neural underpinnings and relation to self-report behaviors. *PLoS One* 6 (6), e21598. <https://doi.org/10.1371/journal.pone.0021598>.
- Arnett, J.J., 2000. Emerging adulthood: a theory of development from the late teens through the twenties. *Am. Psychol.* 55, 469–480.
- Banich, M.T., 2009. Executive function: The search for an integrated account. *Curr. Dir. Psychol. Sci.* 18 (2), 89–94. <https://doi.org/10.1111/j.1467-8721.2009.01615x>.
- Banich, M.T., Milham, M.P., Atchley, R., Cohen, N.J., Webb, A., Wszalek, T., Kramer, A.F., Liang, Z.P., Wright, A., Shenker, J., Magin, R., 2000. fMRI studies of stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *J. Cogn. Neurosci.* 12 (6), 988–1000.
- Banich, M.T., Crowley, T.J., Thompson, L.L., Jacobson, B.L., Liu, X., Raymond, K.M., Claus, E.D., 2007. Brain activation during the stroop task in adolescents with severe substance and conduct problems: a pilot study. *Drug Alcohol Depend.* 90 (2–3), 175–182. <https://doi.org/10.1016/j.drugalcdep.2007.03.009>.
- Bench, C.J., Frith, C.D., Grasby, P.M., Friston, K.J., Paulesu, E., Frackowiak, R.S., Dolan, R.J., 1993. Investigations of the functional anatomy of attention using the stroop test. *Neuropsychologia* 31 (9), 907–922.
- Bickart, K.C., Hollenbeck, M.C., Barrett, L.F., Dickerson, B.C., 2012. Intrinsic amygdala-cortical functional connectivity predicts social network size in humans. *J. Neurosci.* 32 (42), 14729–14741. <https://doi.org/10.1523/JNEUROSCI.1599-12.2012>.
- Casey, B.J., Galvan, A., Somerville, L.H., 2016. Beyond simple models of adolescence to an integrated circuit-based account: a commentary. *Dev. Cogn. Neurosci.* 17, 128–130. <https://doi.org/10.1016/j.dcn.2015.12.006>.
- Carter, C.S., Mintun, M., Cohen, J.D., 1995. Interference and facilitation effects during selective attention: an H215O PET study of stroop task performance. *NeuroImage* 2 (4), 264–272. <https://doi.org/10.1006/nimg.1995.1034>.
- Chorpita, B.F., Tracey, S.A., Brown, T.A., Collica, T.J., Barlow, D.H., 1997. Assessment of worry in children and adolescents: an adaptation of the Penn state worry questionnaire. *Behav. Res. Ther.* 35, 569–581.
- Compton, R.J., Banich, M.T., Mohanty, A., Milham, M.P., Herrington, J., Miller, G.A., Scaif, P.E., Webb, A., Heller, W., 2003. Paying attention to emotion: an fMRI investigation of cognitive and emotional stroop tasks. *Cogn. Affect. Behav. Neurosci.* 3 (2), 81–96.
- Coste, C.P., Sadaghiani, S., Friston, K.J., Kleinschmidt, A., 2011. Ongoing brain activity fluctuations directly account for intertrial and indirectly for intersubject variability in Stroop task performance. *Cereb. Cortex* 21, 2612–2619. <https://doi.org/10.1093/cercor/bhr050>.
- Costello, E.J., Copeland, W., Angold, A., 2011. Trends in psychopathology across the adolescent years: What changes when children become adolescents, and when adolescents become adults? *J. Child Psychol. Psychiatry* 52, 1015–1025. <https://doi.org/10.1111/j.1469-7610.2011.02446.x>.
- Crone, E.A., Dahl, R.E., 2012. Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nat. Rev. Neurosci.* 13 (9), 636–650. <https://doi.org/10.1038/nrn3313>.
- Derryberry, D., Reed, M.A., 2002. Anxiety-related attentional biases and their regulation by attentional control. *J. Abnorm. Psychol.* 111, 225–236.
- Efron, B., 1979. Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7, 1–26.
- Egner, T., Hirsch, J., 2005. Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat. Neurosci.* 8 (12), 1784–1790. <https://doi.org/10.1038/nn1594>.
- Ellis L.K., Rothbart M.K., 2001. Revision of the Early Adolescent Temperament Questionnaire. Poster presented at the 2001 Biennial Meeting of the Society for Research in Child Development, in Minneapolis, Minnesota.
- Ernst, M., 2014. The triadic model perspective for the study of adolescent motivated behavior. *Brain Cogn.* 89, 104–111. <https://doi.org/10.1016/j.bandc.2014.01.006>.
- Gioia, G.A., Isquith, P., Guy, S.C., Kenworthy, L., 2000. Brief rating inventory of executive function. *Child Neuropsychol.* 6 (3), 235–238.
- Grose-Fifer, J., Rodrigues, A., Hoover, S., Zottoli, T., 2013. Attentional capture by emotional faces in adolescence. *Adv. Cogn. Psych.* 9 (2), 81–91. <https://doi.org/10.2478/v10053-008-0134-9>.
- Guyer, A.E., Monk, C.S., McClure-Tone, E.B., Nelson, E.E., Roberson-Nay, R., Adler, A.D., Fromm, S.J., Leibenluft, E., Pine, D.S., Ernst, M., 2008. A developmental examination of amygdala response to facial expressions. *J. Cogn. Neurosci.* 20 (9), 1565–1582. <https://doi.org/10.1162/jocn.2008.20114>.
- Guyer, A.E., Silk, J.S., Nelson, E.E., 2016. The neurobiology of the emotional adolescent: from the inside out. *Neurosci. Biobehav. Rev.* 70, 74–85. <https://doi.org/10.1016/j.neubiorev.2016.07.037>.
- Hankin, B.L., Snyder, H.R., Gulley, L.D., Schweizer, T.H., Bijttebier, P., Nelis, S., et al., 2016. Understanding comorbidity among internalizing problems: integrating latent structural models of psychopathology and risk mechanisms. *Dev. Psychopathol.* 28 (4 pt1), 987–1012. <https://doi.org/10.1017/S0954579416000663>.
- Hankin, B.L., Young, J.F., Abela, J.R.Z., Smolen, A., Jenness, J.L., Gulley, L.D., et al., 2015. Depression from childhood into late adolescence: influence of gender, development, genetic susceptibility, and peer stress. *J. Abnorm. Psychol.* 124 (4), 803–816. <https://doi.org/10.1037/abn0000089>.
- Hankin, B.L., 2015. Depression from childhood through adolescence: risk mechanisms across multiple systems and levels of analysis. *Curr. Opin. Psychol.* 4, 13–20. <https://doi.org/10.1016/j.copsyc.2015.01.003>.
- Hare, T.A., Tottenham, N., Galvan, A., Voss, H.U., Glover, G.H., Casey, B.J., 2008. Biological substrates of emotional reactivity and regulation in adolescence during an emotional go-nogo task. *Biol. Psychiatry* 63 (10), 927–934. <https://doi.org/10.1016/j.biopsych.2008.03.015>.
- Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. *Annu. Rev. Neurosci.* 37, 435–456. <https://doi.org/10.1146/annurev-neuro-062012-170325>.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430. <https://doi.org/10.1126/science.1063736>.
- Hayes, A.F., 2013. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. Guilford Press, N.Y., N.Y.
- Henderson, S.E., Vallejo, A.I., Ely, B.A., Kang, G., Krain Roy, A., Pine, D.S., Stern, E.R., Gabbay, V., 2014. The neural correlates of emotional face-processing in adolescent depression: a dimensional approach focusing on anhedonia and illness severity. *Psychiatry Res.* 224 (3), 234–241. <https://doi.org/10.1016/j.psychres.2014.09.006>.
- Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. *P Natl. Acad. Sci. USA* 96 (16), 9379–9384.
- Ishai, A., Ungerleider, L.G., Haxby, J.V., 2000. Distributed neural systems for the generation of visual images. *Neuron* 28 (3), 979–990.
- Kaiser, R.H., Andrews-Hanna, J.R., Spielberg, J.M., Warren, S.L., Sutton, B.P., Miller, G.A., et al., 2015. Distracted and down: neural mechanisms of affective interference in subclinical depression. *Soc. Cogn. Affect. Neurosci.* 10 (5), 654–663. <https://doi.org/10.1093/scan/nsu100>.
- Kane, M.J., Engle, R.W., 2003. Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *J. Exp. Psychol. Gen.* 132 (1), 47–70.
- Killgore, W.D.S., Gruber, S.A., Yurgelun-Todd, D.A., 2007. Depressed mood and lateralized prefrontal activity during a Stroop task in adolescent children. *Neurosci. Lett.* 416 (1), 43–48. <https://doi.org/10.1016/j.neulet.2007.01.081>.
- Kovacs, M., 1992. Children's Depression Inventory. Multi-Health Systems, Inc., North Tonawanda, NY.
- Laeger, I., Dobel, C., Dannlowski, U., Kugel, H., Grotegerd, D., Kissler, J., et al., 2012. Amygdala responsiveness to emotional words is modulated by subclinical anxiety and depression. *Behav. Brain Res.* 233 (2), 508–516. <https://doi.org/10.1016/j.bbr.2012.05.036>.

- Lewis-Peacock, J.A., Norman, K.A., 2014. Multivoxel pattern analysis of functional MRI data. In: Gazzaniga, M.S., Mangun, G.R. (Eds.), *The Cognitive Neurosciences*. MIT Press, Cambridge, pp. 911–920.
- Lewis-Peacock, J.A., Postle, B.R., 2012. Decoding the internal focus of attention. *Neuropsychologia* 50, 470–478. <https://doi.org/10.1016/j.neuropsychologia.2011.11.006>.
- MacLeod, C., MacDonald, P., 2000. Interdimensional interference in the stroop effect: uncovering the cognitive and neural anatomy of attention. *Trends Cogn. Sci.* 4 (10), 383–391.
- March, J.S., Parker, J.D., Sullivan, K., Stallings, P., Conners, C.K., 1997. The multi-dimensional anxiety scale for children (MASC): factor structure, reliability, and validity. *J. Am. Acad. Child Psychiatry* 36 (4), 554–565. <https://doi.org/10.1097/00004583-199704000-00019>.
- Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>.
- Mills, K.L., Goddings, A.-L., Herting, M.M., Meuwese, R., Blakemore, S.-J., Crone, E.A., et al., 2016. Structural brain development between childhood and adulthood: convergence across four longitudinal samples. *NeuroImage* 141, 273–281. <https://doi.org/10.1016/j.neuroimage.2016.07.044>.
- Monk, C., McClure, E., Nelson, E., Zarahn, E., Bilder, R.M., Leibenluft, E., et al., 2003. Adolescent immaturity in attention-related brain engagement to emotional facial expressions. *NeuroImage* 20, 420–428.
- Murray, R.J., Brosch, T., Sander, D., 2014. The functional profile of the human amygdala in affective processing: insights from intracranial recordings. *Cortex* 60, 10–33. <https://doi.org/10.1016/j.cortex.2014.06.010>.
- Niendam, T.A., Laird, A.R., Ray, K.L., Dean, Y.M., Glahn, D.C., Carter, C.S., 2012. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn. Affect. Behav. Neurosci.* 12 (2), 241–268. <https://doi.org/10.3758/s13415-011-0083-5>.
- Pardo, J.V., Pardo, P.J., Janer, K.W., Raichle, M.E., 1990. The anterior cingulate cortex mediates processing selection in the stroop attentional conflict paradigm. *Proc. Natl. Acad. Sci. USA* 87 (1), 256–259.
- Paus, T., Keshavan, M., Giedd, J.N., 2008. Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* 9 (12), 947–957. <https://doi.org/10.1038/nrn2513>.
- Phelps, E.A., 2005. Emotion and cognition: insights from studies of the human amygdala. *Annu. Rev. Psychol.* 57, 27–53. <https://doi.org/10.1146/annurev.psych.56.091103.070234>.
- Polk, T.A., Drake, R.M., Jonides, J.J., Smith, M.R., Smith, E.E., 2008. Attention enhances the neural processing of relevant features and suppresses the processing of irrelevant features in humans: a functional magnetic resonance imaging study of the stroop task. *Neuroscience* 28, 13786–13792. <https://doi.org/10.1523/JNEUROSCI.1026-08.2008>.
- Pourtois, G., Schettino, A., Vuilleumier, P., 2013. Brain mechanisms for emotional influences on perception and attention: what is magic and what is not. *Biol. Psychol.* 92 (3), 492–512. <https://doi.org/10.1016/j.biopsycho.2012.02.007>.
- Purmann, S., Pollmann, S., 2015. Adaptation to recent conflict in the classical color-word Stroop-task mainly involves facilitation of processing of task-relevant information. *Front. Human Neurosci.* 9, 88. <https://doi.org/10.3389/fnhum.2015.00088>.
- Shulman, E.P., Smith, A.R., Silva, K., Icenogle, G., Duell, N., Chein, J., Steinberg, L., 2016. The dual systems model: review, reappraisal, and reaffirmation. *Dev. Cogn. Neurosci.* 17, 103–117. <https://doi.org/10.1016/j.dcn.2015.12.010>.
- Snyder, H.R., Gulley, L.D., Bijttebier, P., Hartman, C.A., Oldehinkel, A.J., Mezulis, A., et al., 2015. Adolescent emotionality and effortful control: core latent constructs and links to psychopathology and functioning. *J. Pers. Soc. Psychol.* 109 (6), 1132–1149. <https://doi.org/10.1037/pspp0000047>.
- Steinberg, L., 2010. A dual systems model of adolescent risk-taking. *Dev. Psychobiol.* 52, 216–224. <https://doi.org/10.1002/dev.20445>.
- Steinberg, L., Cauffman, E., Woolard, J., Graham, S., Banich, M., 2009. Are adolescents less mature than adults?: minors' access to abortion, the juvenile death penalty, and the alleged APA "flip-flop". *Am. Psychol.* 64 (7), 583–594. <https://doi.org/10.1037/a0014763>.
- Steinhauser, M., Flaisch, T., Meinzer, M., Schupp, H.T., 2016. Cognitive control modulates preferential sensory processing of affective stimuli. *Neuropsychologia* 91, 435–443. <https://doi.org/10.1016/j.neuropsychologia.2016.09.009>.
- Tackett, J.L., Lahey, B.B., van Hulle, C., Waldman, I., Krueger, R.F., Rathouz, P.J., 2013. Common genetic influences on negative emotionality and a general psychopathology factor in childhood and adolescence. *J. Abnorm. Psychol.* 122 (4), 1142–1153. <https://doi.org/10.1037/a0034151>.
- Tellegen, A., Watson, D., Clark, L.A., 1999. On the dimensional and hierarchical structure of affect. *Psychol. Sci.* 10 (4), 297–303.
- Tottenham, N., Tanaka, J.W., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., Nelson, C., 2009. The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res.* 168 (3), 242–249. <https://doi.org/10.1016/j.psychres.2008.05.006>.
- Veroude, K., Jolles, J., Croiset, G., Krabbendam, L., 2013. Changes in neural mechanisms of cognitive control during the transition from late adolescence to young adulthood. *Dev. Cogn. Neurosci.* 5, 63–70. <https://doi.org/10.1016/j.dcn.2012.12.002>.
- Vijayakumar, N., Whittle, S., Yücel, M., Dennison, M., Simmons, J., Allen, N.B., 2014. Prefrontal structural correlates of cognitive control during adolescent development: a 4-year longitudinal study. *J. Cogn. Neurosci.* 26 (5), 1118–1130. https://doi.org/10.1162/jocn_a_00549.
- Vuilleumier, P., 2015. Affective and motivational control of vision. *Curr. Opin. Neurol.* 28 (1), 29–35. <https://doi.org/10.1097/WCO.0000000000000159>.
- Wager, T.D., Waugh, C.E., Lindquist, M., Noll, D.C., Fredrickson, B.L., Taylor, S.F., 2009. Brain mediators of cardiovascular responses to social threat, Part I: reciprocal dorsal and ventral sub-regions of the medial prefrontal cortex and heart-rate reactivity. *NeuroImage* 47, 821–835. <https://doi.org/10.1016/j.neuroimage.2009.05.043>.
- Wechsler, D., 1999. *Wechsler Abbreviated Scale of Intelligence*. Pearson, Upper Saddle River, NJ.
- Woolgar, A., Hampshire, A., Thompson, R., Duncan, J., 2011. Adaptive coding of task-relevant information in human frontoparietal cortex. *J. Neurosci.* 31 (41), 14592–14599. <https://doi.org/10.1523/jneurosci.2616-11.2011>.
- Zanto, T.P., Rubens, M.T., Thangavel, A., Gazzaley, A., 2011. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat. Neurosci.* 14 (5), 656–661. <https://doi.org/10.1038/nn.2773>.
- Zanto, T.P., Rubens, M.T., Bollinger, J., Gazzaley, A., 2010. Top-down modulation of visual feature processing: the role of the inferior frontal junction. *NeuroImage* 53 (2), 736–745. <https://doi.org/10.1016/j.neuroimage.2010.06.012>.