Composite Measures of Brain Activation Predict Individual Differences in Behavioral Stroop Interference

Louisa L. Smith[1], Hannah R. Snyder[2], Benjamin L. Hankin[3], and Marie T. Banich[1]

Abstract

■ The goal of the current study was to interrogate aspects of the cascade-of-control model [Banich, M. T. Executive function: The search for an integrated account. *Current Directions in Psychological Science*, *18*, 89–94, 2009; Banich, M. T. The Stroop effect occurs at multiple points along a cascade of control: Evidence from cognitive neuroscience approaches. *Frontiers in Psychology*, *10*, 2164, 2019], a neurocognitive model that posits how portions of pFC interact in a cascade-like manner to overcome interference from task-irrelevant information, and to test whether it could be used to predict individual differences in cognitive control outside the scanner. Participants (n = 62) completed two fMRI Word–Picture Stroop tasks, one containing emotional stimuli and one containing non-emotional stimuli, as well as a behavioral out-of-scanner Color–Word Stroop task at each of two time points. In a departure from the traditional approach of using a single task contrast to index neural activation across all ROIs, the current study utilized specific ROI by contrast pairings selected based on the specific level of control hypothesized by the cascade-of-control model to occur within that region. In addition, data across both tasks and both time points were combined to create composite measures of neural activation and of behavior. Consistent with the cascade-of-control model, individual differences in brain activation for specific contrasts within each of the three ROIs were associated with behavioral interference on the standard Color–Word Stroop task. Testing of alternative models revealed that these brain–behavior relationships were specific to the theoretically driven ROI by contrast pairings. Furthermore, such relationships were not observed across single-task and single-time point measures, but instead emerged from the composite measures. These findings provide evidence that brain activation observed across multiple regions of frontal cortex, each of which likely exerts cognitive control in a differential manner, is capable of predicting individual differences in behavioral performance. ■

INTRODUCTION

Cognitive control allows an individual to prioritize and act on information that is goal-relevant. Such goal-directed behavior is implemented by neural activation in distributed regions and networks throughout the brain, which serve to bias toward task-relevant processing and protect against interference from competing information. The goal of the current study is to test predictions made by the cascade-of-control model (Banich, 2009, 2019), a neurocognitive model that posits that cognitive control is enacted via multiple regions of frontal cortex that each resolve unique levels of interference in a cascade-like manner, and to utilize this model to predict individual differences in behavioral performance outside of the magnet. Specifically, this model asserts that different regions of pFC play somewhat distinct roles in exerting control, and that the degree to which a task set is effectively implemented by lateral prefrontal regions determines the amount of control that must be enacted at later stages along the cascade. Using this model as a guiding theoretical framework, we test the ability of multitask composite measures of neural activation, selected on the basis of the cascade-of-control model, to predict individual differences in out-of-scanner behavioral performance. We first present the tenets of the cascade-of-control model and then discuss its use to examine brain–behavior relationships.

Cascade-of-control Model

Substantial research has identified regions of both lateral and medial pFC as playing a critical role in cognitive control (Derrfuss, Brass, Neumann, & von Cramon, 2005). One model designed to explicitly outline the neural systems required for Stroop task performance is the cascade-of-control model (Banich, 2009, 2019). In the classic Stroop task (Stroop, 1935), participants are presented with words printed in different ink colors and must respond based on the ink color in which a word is written, rather than on the meaning of the word. This is more difficult and requires more control on incongruent trials, in which the word conflicts with the ink color, (e.g., "RED" printed in blue ink) as compared with congruent trials, in which no conflict exists ("RED" printed in red ink). As discussed in Banich (2019), conflict can arise at many levels in the Stroop task: between the more automatic, but task-irrelevant, process of word reading as compared with ink-color identification; between the meaning of

[1]University of Colorado Boulder, [2]Brandeis University, Waltham, MA, [3]University of Illinois–Urbana Champaign

Journal of Cognitive Neuroscience 35:5, pp. 781–801
https://doi.org/10.1162/jocn_a_01977

"red" as compared with the meaning of "blue"; and between responses linked to red as opposed to blue. The cascade-of-control model posits that specific prefrontal regions play more prominent roles than others in resolving conflict at these different levels. In particular, the model proposes that abstract task set implementation occurs within inferior frontal junction, biased selection among working memory representations within mid-dorsolateral pFC (DLPFC), and late-stage response selection within mid-cingulate regions.

Furthermore, the cascade-of-control model aims to provide an integrative account of how these regions interact in a cascade-like manner to overcome interference from task-irrelevant information. Central to this model is the idea that control is implemented at multiple levels of interference that span stimulus inputs to motor outputs. Furthermore, each level of interference is resolved by a different cognitive operation that is enacted, to a degree, by an individual brain region. Critically, ineffective control enacted at early stages within the cascade may be compensated for by downstream brain regions, effectively sparing behavioral outcomes. Stated more concretely, the behavior of individuals who maintain a task set poorly may only suffer if downstream regions are unable to "pick up the slack." As such, it is important to consider the role of multiple brain regions in predicting behavior, rather than a single brain–behavior relationship in isolation.

In a departure from the traditional approach of using a single condition contrast (e.g., incongruent vs. congruent trials) to index neural activation across multiple ROIs regardless of their proposed role in resolving interference, we determine the appropriate task contrast for each individual ROI on the basis of the cascade-of-control model. This novel approach of using unique ROI by contrast pairings is driven by the assertion that the nature of Stroop interference differs across levels of control and allows us to capture the activation within each ROI that is relevant to the specific cognitive process hypothesized to occur within that region. In the following section, we discuss the role of each ROI in implementing control during a standard Color–Word Stroop task. We describe the specific level of control that is engendered within each ROI as posited by the cascade-of-control model (Banich, 2009, 2019) and provide the rationale for the chosen ROI by contrast pairing. A discussion of the evidence that these regions play the specific role ascribed to them is beyond the scope of the current introduction but can be found in detail in Banich (2019).

Inferior Frontal Junction: Task Set Implementation

First along the cascade is inferior frontal junction (IFJ), a region of lateral pFC that engenders control by implementing a task set that serves to bias processing toward task-relevant and/or away from task-irrelevant processing in sensory and perceptual regions of posterior cortex. In the context of the Color–Word Stroop task, IFJ would implement up-regulation of color-related processing and/or down-regulation of word-related processing. Such biasing occurs regardless of the current trial type and so is present during incongruent (INC) and congruent (CON) trials, as well as the neutral (NEU) trials also used in the

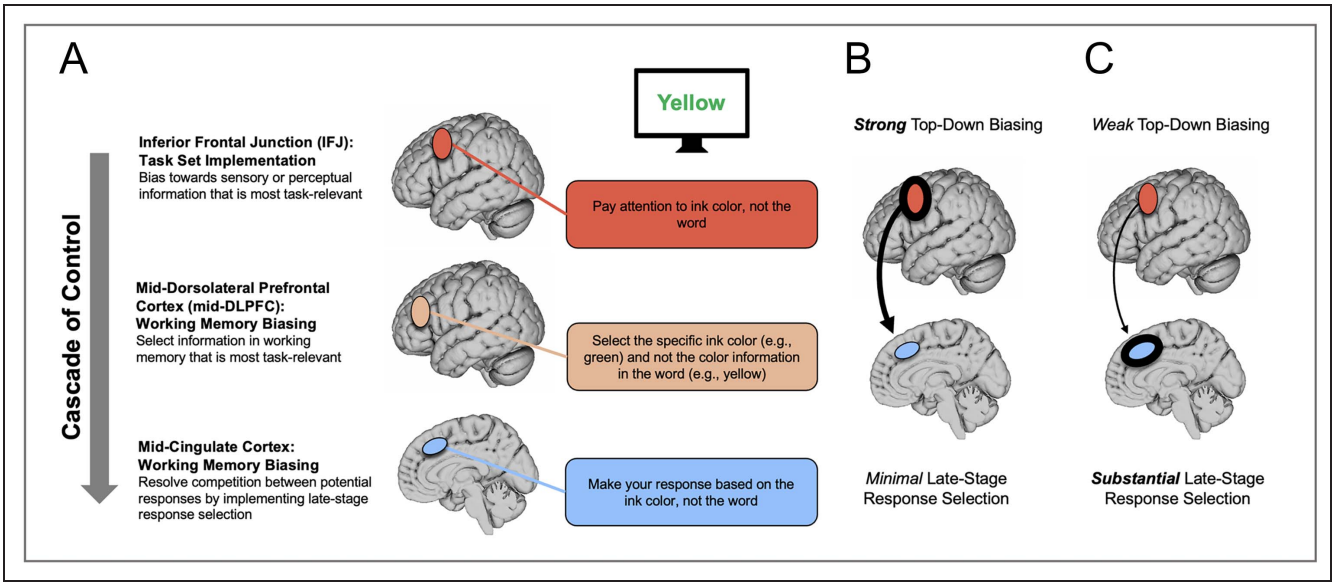

Figure 1. (A) The cascade-of-control model (Banich, 2009, 2019) outlining the brain regions that are involved in controlling interference in the Stroop task. For an incongruent trial, such as the word "Yellow" written in green ink, control is implemented via a cascade. First, regions of IFJ enact an abstract task set that serves to bias sensory and perceptual processing toward the task-relevant color information and/or away from the task-irrelevant word information. Next along the cascade, mid-DLPFC regions select the task-relevant information that should be maintained in working memory (e.g., green not yellow). At the final stage considered within the current study, caudal regions of mid-cingulate bias late-stage response selection toward the task-appropriate response (e.g., button press for green). Importantly, the degree to which one region is active in controlling Stroop interference depends on how well control has been implemented at prior points in the cascade. (B) Example of strong top–down biasing via IFJ. (C) Example of weak top–down biasing via IFJ.

current study, which in the context of the Color–Word Stroop task would be trials in which the word is not related to color. In the context of the current study, we index relevant activation by contrasting all three trial types (incongruent, congruent, and neutral) with fixation (Inc + Con + Neu > Fix). This logic is consistent with others who have argued that any regions involved in setting a task set should be consistently observed across task conditions (e.g., Dosenbach et al., 2006). Although it is true that other regions may be activated by this contrast (e.g., perceptual processing regions), we highlight the role of IFJ in implementing a task set based on our prior work (e.g., Banich et al., 2000) and other findings that this region plays a role in top–down biasing (Baldauf & Desimone, 2014; Muhle-Karbe, Andres, & Brass, 2014; Brass, Derrfuss, Forstmann, & von Cramon, 2005). (For other viewpoint on other brain regions that may play a major role in top–down biasing, see Dosenbach et al., 2006.)

Mid-DLPFC: Working Memory Biasing

Next along the cascade, mid-DLPFC is posited to bias selection toward the task-relevant information in working memory, which in the case of the Color–Word Stroop task would be the identity of the task-relevant ink color. Such biasing is not required during neutral trials (e.g., "TABLE" written in red ink), as the task-irrelevant word identity does not contain task-relevant color information. In contrast, task-relevant information is present in the task-irrelevant dimension during both incongruent and congruent trials, because in both cases the word names a color. Thus, neutral trials are contrasted with both incongruent and congruent trials to isolate relevant activation in mid-DLPFC (Inc + Con > Neu).

Mid-cingulate: Late-stage Response Selection

Finally, late-stage response-related selection is posited to be implemented by mid-cingulate, which in the case of the Color–Word Stroop task would bias processing toward the motor response associated with the task-relevant word. Relevant activation is defined by the contrast between incongruent trials in which the task-relevant and task-irrelevant information map to different responses, and congruent trials in which both dimensions map to the same response. This contrast likely also captures additional, non-response-related interference including semantic interference. However, prior work has demonstrated that effects within mid-cingulate are specific to incongruent trials in which the task-irrelevant information is response-eligible, suggesting that this interference is primarily driven by motor mappings and not semantics (Milham & Banich, 2005; Milham et al., 2001).

The cascading nature of this system dictates that the successful enactment of an appropriate task set by IFJ will determine the degree of control required during late-stage selection in mid-cingulate. Consistent with this model, in a combined fMRI and ERP Stroop task, Silton et al. (2010) found that the relationship between early IFJ activity (300–400-msec post-stimulus presentation) and behavioral performance was mediated by late ACC activity (520–680 msec). Furthermore, a significant interaction between IFJ and ACC activity suggested that ACC mainly influences behavioral performance only when IFJ is not optimally engaged. Alternative models, such as those in which early ACC activity was assumed to influence later IFJ activity, could not predict behavioral performance. In the next section, we discuss the promise and pitfalls of examining brain–behavior relationships within the context of functional neuroimaging studies.

Individual Differences and Brain Activation

fMRI offers a potentially powerful tool to link brain activation to behavioral outcomes. fMRI has been used extensively to identify brain regions and networks that underlie a wide range of cognitive processes including those that support cognitive control. Brain regions are assumed to support a particular mental process when, across a group of participants, activation is higher during engagement in a task condition that requires that process as compared with a control condition that does not (e.g., conflict vs. no conflict). For a region to be identified through this approach, activation differences must be consistently observed across participants, such that the average difference is reliably distinguishable from zero. There are two limitations, however, with these group averaged data. First, they do not identify which regions may contribute most directly to the accuracy or speed of a particular mental process that is being examined. Nor, because they are group averages, do they provide insights into how activation of such regions may influence the ultimate behavioral performance on the task being performed, nor how individual differences in behavioral characteristics (e.g., level of cognitive control ability) may relate to variations in brain activation.

These issues are important because research indicates that there are stable individual differences across a wide range of cognitive functions including executive functions such as inhibition, working memory, and task shifting (e.g., Friedman et al., 2016), as well as long-term memory (e.g., Unsworth, 2019), language abilities (e.g., Kidd, Donnelly, & Christiansen, 2018), and attentional processing (e.g., Chechlacz, Gillebert, Vangkilde, Petersen, & Humphreys, 2015). For example, Friedman et al. (2008, 2016) used structural equation modeling to identify a Common Executive Function (Common EF) latent factor that is hypothesized to reflect the ability to maintain and manage task goals as well as bias processing in accordance with those goals. In an analysis of EF stability in 840 individuals across a 6-year period, Friedman et al. (2016) found a cross-wave correlation of .86 for the latent Common EF factor and a cross-wave correlation of .45

more specifically for Stroop interference, suggesting that individual differences in Common EF are stable across time. Suggesting that this factor is linked to real-world performance, Common EF has been linked to self-reported EF, including the ability to manage and retrieve goals (Gustavson, Miyake, Hewitt, & Friedman, 2015).

Such findings have prompted a search for the neural underpinnings of observed individual differences because, in part, understanding the neural underpinnings of individual differences in cognitive control is a crucial component of better predicting, diagnosing, treating, and ultimately preventing brain disorders (Elliott, Knodt, & Hariri, 2021). The typical approach used to investigate such relationships is to examine the covariation between the magnitude of differences in brain activation between a critical contrast of interest and the magnitude of a behavioral outcome of interest. In the case of the Stroop task, behavioral performance is typically measured by a RT interference score that is calculated as the difference in RT on the control-demanding incongruent trials compared with the less demanding congruent trials. Brain regions that underlie individual differences in Stroop interference would thus be identified as those in which activation differences are significantly associated with the size of the RT interference.

In this approach, the individual differences variable (e.g., RT interference score, level of Common EF ability) is typically included as a covariate in the general linear model (GLM) analysis. However, such an approach has been criticized as likely to lead to false positive findings (Vul, Harris, Winkielman, & Pashler, 2009) and may require very large sample sizes (> 1000; Marek et al., 2022). However, others have argued that with reliable single time point measures of brain and behavior, samples sizes in the range of about 80 individuals are adequate for detecting brain–behavior relationships (Grady, Rieck, Nichol, Rodrigue, & Kennedy, 2021). Moreover, whereas the focus with regard to statistical power had generally focused on the number of participants in a study (n), which influences between-subjects variance, recent discussions of brain–behavior relationships in neuroimaging studies have noted that statistical power is equally influenced by the robustness of the measures obtained on each individual, which influences within-subject variance (Chen et al., 2022; Baker et al., 2021) and that equivalent power can be obtained by different combinations or pairings of between- and within-subject variance. One example of an attempt to reduce within-subject variance are studies employing the Midnight Scan Club data set (Gordon et al., 2017), which involve 5 hr of rest data and 6 hr of task data on a single individual to obtain stable individual differences measures of both brain and behavior.

In the current study, we have considered these two paths to achieving greater reliability in brain–behavior correlations (Gratton, Nelson, & Gordon, 2022)—a larger N or deeper phenotyping to obtain a more stable within-subject measure—and take an intermediate approach. This intermediate approach involved obtaining measures both of brain and behavior at two distinct time points assessed 2 years apart and averaging across them, while also obtaining measures of brain activation across more than one task, as discussed in more detail below.

There are several advantages to using composite measures composed of data collected across multiple tasks and multiple time points. By varying task-related attributes such as stimulus category and valence across the two fMRI paradigms, we reduce the degree to which the composite measures may reflect stimulus-specific processing. Similarly, by combining measures that were collected at two different time points, we reduce the likelihood that the composite measures reflect state-specific factors such as alertness, which is known to impact cognitive processes (e.g., Killgore, 2010). In addition, composite measures of brain activation have been shown to have higher reliability than single task measures (e.g., Sheu, Jennings, & Gianaros, 2012).

Another issue with regard to assessing brain–behavior relationships is that it is not always obvious what particular contrast or comparison of conditions should be used to assess brain activation. Typically, a particular contrast is employed for all brain areas under investigation, such as a contrast in activation between incongruent and congruent trials in the Stroop task, and then a whole-brain GLM for this contrast with the covariate of an individual differences variable (e.g., level of executive function) is performed. Yet, if one has reason to believe that different brain regions may be more involved in particular subprocesses of a given task, then such an approach is not ideal. Rather, it may be more fruitful to utilize different contrasts for different brain regions to examine brain–behavior relationships. To do so requires a theoretical model of the predominant role of each major brain region involved in a task. In the current study, we take such an approach to guide our investigation using the cascade-of-control framework.

Current Study

The current study investigates whether individual differences in behavioral Stroop interference on a standard Color–Word Stroop task given outside the magnet are predicted by composite measures of task-related activation in three regions of frontal cortex: IFJ, mid-DLPFC, and mid-cingulate during performance of two Word–Image Stroop tasks. We evaluate the brain–behavior relationships predicted by the cascade-of-control model, focusing specifically on the theoretically derived ROI by contrast pairings that reflect the cognitive processes hypothesized to occur within each region. To do so, we test the following theoretically motivated hypotheses: (1) Significant associations will be observed between behavioral interference and activation within each ROI. (2) Because of the nature of the control that is differentially engaged for each ROI, associations between ROI

activation and behavioral interference will be contrast-specific. (3) Consistent with the findings of Silton et al. (2010), we expect to observe a significant interaction between activation in IFJ and mid-cingulate, such that the relationship between mid-cingulate activation and behavioral performance is strongest when IFJ activation is low. We additionally hypothesize that this interaction will also be contrast-specific.

We tested these hypotheses in a sample of 62 adult women who participated in both fMRI and behavioral testing sessions as part of a larger longitudinal study. Although smaller than the number generally recommended to reach ideal levels of effect stability for a single time-point observation (e.g., Grady et al., 2021; Yarkoni, 2009), this sample size is substantially larger than is typical of fMRI studies (Poldrack et al., 2017). At each of two time points separated by approximately 2 years, participants completed two fMRI Word–Picture Stroop tasks—one consisting of affectively neutral stimuli, and the other emotional stimuli—as well as a behavioral Color–Word Stroop task. We then use composite measures comprising all available data for a given individual to index brain activation and behavior (see Methods section).

To the degree that meaningful individual differences exist within functional brain activation during the resolution of interference, these differences should predict behavioral performance outside of those tasks during which the activation was measured. This approach is useful because measuring activation and performance during the same task may produce relationships between brain activation and performance that are driven by time on task (Grinband et al., 2011) or the same type of stimulus- or state-specific processing as previously mentioned. To this end, we chose to index behavioral interference using a separate Stroop task that was administered on a different day. We sought to minimize the likelihood that observed brain–behavior relationships might be driven by task-related attributes by selecting a behavioral task that indexes cognitive control, the standard Color–Word Stroop task, and then deriving brain activation from distinct but related word-image Stroop tasks. One important distinction between the tasks is that word reading serves as the task-relevant process/dimension in the fMRI tasks but as the task-irrelevant process/dimension in the behavioral task. As such, any observed relationships cannot be driven by commonalities between the nature of the task-relevant information or conversely the task-irrelevant information. The tasks also differed substantially in that word reading was paired with task-irrelevant images in the fMRI tasks and with task-relevant ink color in the behavioral task. Finally, the tasks differed with respect to response type, as the fMRI tasks required finger movement whereas the behavioral task required a verbal response. As such, the metrics used to examine cognitive control behaviorally and in the magnet are distinct, making it unlikely that any association observed results from stimulus or response similarity.

In summary, the current study expands upon prior work in several important ways. First, although the cascade-of-control model describes the role of multiple brain regions, the current study is the first to simultaneously consider three ROIs and their association with behavior. Second, this is one of very few studies that utilize a theoretical neurocognitive model to identify and test contrast-specific brain–behavior relationships, as compared with the traditional approach of using a single contrast across ROIs (e.g., Tang, Etzel, Kizhner, & Braver, 2021). Finally, we use a number of approaches to increase the reliability and stability of our indices of individual differences related to brain and behavior.

METHODS

Participants

Data from unrelated adult women were collected as part of the Colorado Cognitive Neuroimaging Family Emotion Research (CoNiFER) study, a two-time-point longitudinal assessment of adolescent children and their parents. Only female parents were included in the present analyses as an insufficient number of male parents ($n = 5$) chose to participate. All participants had previously taken part in the Genes, Environment and Mood study (R01MH077195, Hankin, P.I.) and an associated follow-up study (R21MH102210, Hankin, P.I.). Families were drawn from an unselected community sample, recruited from the Denver metro area via public schools and using direct mail to target zip codes to maximize demographic and socioeconomic diversity. For details of the two samples and studies, see Hankin et al. (2015) and Snyder, Friedman, and Hankin (2019).

Participants were free of history of neurological insult, were fluent English speakers, and did not report having dyslexia or difficulty reading. Informed consent was obtained from all participants, and all procedures were approved by the University of Colorado institutional review board. All inclusion/exclusion criteria were established before data analysis. fMRI data were subject to the following exclusion criteria: Low accuracy (accuracy below 60% on incongruent, congruent, or neutral trials), excessive signal dropout in frontal or temporal regions, excessive motion (greater than 3 mm), or failure to meet multiple criteria.

Time Point 1 (TP1): 70 women participated at TP1. Six participants were excluded from both fMRI tasks because of low accuracy ($n = 2$), excessive signal dropout ($n = 2$), or failure to meet multiple criteria ($n = 2$). No participants were excluded due to excessive motion. An additional two participants did not complete the Color–Word Stroop task and were excluded from analysis. Data from six participants were included for the emotional task but excluded from the non-emotional task because low accuracy ($n = 4$), or excessive movement ($n = 2$). Data from three participants were included for the non-emotional

task but excluded from the emotional task because of excessive movement greater than 3 mm ($n = 3$). No data were excluded due to low accuracy, excessive signal dropout, or failure to meet multiple criteria. These exclusions resulted in usable data from 56 participants in the non-emotional task, 59 participants in the emotional task, and usable data across both tasks for 53 participants.

Time Point 2 (TP2): 55 women participated at TP2. Five participants were excluded from both fMRI tasks because of low accuracy ($n = 1$), excessive signal dropout ($n = 1$), excessive movement, or failure to meet multiple criteria ($n = 1$). Data from 10 participants were included for the emotional task but excluded from the non-emotional task because of low accuracy ($n = 5$), excessive movement ($n = 4$), or failure to complete the task ($n = 1$), No data were excluded due to excessive signal dropout or failure to meet multiple criteria. Data from two participants were included for the non-emotional task but excluded from the emotional task because of low accuracy ($n = 1$), or excessive movement ($n = 1$). No participants were excluded due to excessive signal dropout or failure to meet multiple criteria. These exclusions resulted in usable data from 40 participants in the non-emotional task and 48 participants in the emotional task, and usable data across both tasks for 38 participants.

Sixty-seven participants had usable data for at least one of four fMRI observations (two time points × two tasks) and data from at least one Color–Word Stroop task session. Of these participants, 32 had usable data at all four fMRI observations, 10 had usable data at three observations, and 20 had usable data at two observations. Five had usable data for only one observation. Ideally, each participant would have usable data at all four observation points (two tasks × two time points). However, to maintain a sufficient sample size, all participants with usable data for at least two fMRI tasks and at least one behavioral task were included in the analysis. Thus, the final sample consisted of 62 participants. Of these participants, 45 (73%) had usable data from at least one fMRI task at both TP1 and TP2, and 61 (98%) had usable data from at least one time point for both the emotion and non-emotional tasks. The average age across both time points ranged from 36.5 years to 65.6 years ($M = 49.5$ years, $SD = 6.2$ years).

A power analysis employing G*Power 3.1.9.4 (Faul, Erdfelder, Lang, & Buchner, 2007) estimated an achieved power of $1 - \beta = .85$ ($\alpha = .05$) to detect a small to medium effect size ($f^2 = 0.15$) for each regression parameter in the primary study analysis. Although larger sample sizes have been recommended to reach ideal levels of effect stability in brain–behavior relationships (e.g., Grady et al., 2021; Yarkoni, 2009), the multitask composite measure approach utilized in the current study required a compromise between the number of participants and the quantity of data collected from each participant.

Procedure

At each of two time points separated by approximately 2 years, participants completed a neuroimaging session as well as a behavioral session. The neuroimaging and behavioral sessions typically occurred within 2 months of one another, but, in rare cases, the sessions were separated by up to 6 months. The neuroimaging sessions consisted of three fMRI tasks, including both an emotional and a non-emotional Stroop task, as well as a verb generation task that is not of interest for the current report. The non-emotional Stroop task always preceded the emotional task so as to not confer emotional meaning to neutral face stimuli used in the non-emotional task. Resting state and spectroscopy data were also collected but are not included in the present analyses. Behavioral testing sessions included a battery of questionnaires and cognitive tasks. These included a Color–Word Stroop task, which is the behavioral outcome of interest in the current study, as well as a choice RT task, which is used as a control variable.

Measures

Neuroimaging Tasks

Emotional word–emotional face Stroop task. The goal of this task was to identify those brain regions that are activated when cognitive control must be exerted in the context of semantic and response conflict between two categories of emotional items. Individuals made a semantic decision about whether words were positively valenced (i.e., "cheerful," "joy," "happy," "delighted") or negatively valenced (i.e., "furious," "mad," "angry," "rage"), and indicated their response with either a right or left button press. The words were superimposed on three types of pictures: faces with a happy expression, faces with an angry expression, and cars. Stimuli used in the present study were designed to be identical to those used with the adolescent offspring of the participants (which is the focus of another report). As part of the goal in the adolescent sample was to examine the development of prefrontal-amygdala connectivity as assessed by fMRI, angry and happy faces were used as they show good reliability in this regard (Haller et al., 2018). Cars, rather than neutral faces, were used as neutral stimuli because our prior work suggests that youth in the age range we tested show highly similar responses to faces with neutral and emotional expressions (Banich et al., 2019).

Faces were drawn from the NIMSTIM database (Tottenham et al., 2009), using a set of 24 posers, with two angry expression and two happy expression images drawn from each poser. Each poser was seen four times, twice with an angry and twice with a happy expression. Thirty-two car images were drawn from Herzmann and Curran (2011). For half of the trials, the word was placed in an upper position with regard to the object, and on half the trials, it was placed in a lower position with regard to the object.

Varying the position of the word made it less likely that an individual could narrow their focus of attention to just the central portion of the display where the word would appear so as to filter out the background image.

Three types of trials were constructed. On congruent trials, task-relevant words were drawn from the same semantic category (e.g., a positively valenced word superimposed on a happy face). On incongruent trials, the task-relevant word was drawn from the opposite semantic category than the task-irrelevant face (e.g., a negatively valenced word superimposed on a picture of a happy face; a positively valenced word superimposed on an angry face). On neutral trials, the task-irrelevant object was a car, which has no semantic or response relationship to either positively valenced or negatively valenced words. Each stimulus was displayed for a fixed period of 1380 msec (3 repetition times [TRs]), followed by a fixed intertrial interval of 480 msec (1 TR). Sample stimuli are shown in Figure 2.

Similar to designs used previously in our laboratory (Andrews-Hanna et al., 2011; Banich et al., 2009), this study was a mixed blocked/event-related design. Blocks consisted of trials that were, on average, two thirds specific to that block (e.g., incongruent) and one third a set of neutral trials that were common across blocks, referred to as neutral frequent trials. In the neutral block, two thirds of trials were a set of neutral infrequent trials, and one third the same neutral trials that were common across blocks (i.e., neutral frequent). Six triads of blocks were created (e.g., INC, CIN), and these were flanked in a run by fixation blocks (Fix). Fixation blocks lasted for 23.04 sec (48 TRs) each and served as the baseline for the contrasts discussed below. Each run consisted of three triads, so, for example, a run might consist of FINCFNICFCINF. Two orders of blocks were employed, with button responses counterbalanced across participants.

The exact number of trials per condition differed slightly. For half of the participants, one of the congruent blocks contained two-third frequent neutral trials and one-third congruent trials, whereas the other half of participants were presented with one of the incongruent blocks consisting of two-third frequent neutral trials and one-third incongruent trials. As a result, all participants received 216 trials, with half of the participants seeing 44 incongruent and 48 congruent trials, and the other half seeing 48 incongruent and 44 congruent trials, with all participants seeing 48 neutral infrequent and 76 neutral frequent trials.

Neuroimaging neutral word–neutral image Stroop task. The goal of this task was to identify those brain regions that are activated when cognitive control must be exerted in the face of semantic and response conflict between two categories of non-emotional items. This task was identical to the emotional task (described above), except that the stimuli were non-emotional. Individuals made a semantic decision about whether words were face-related (i.e., "hair," "nose," "chin," "eyebrow") or house-related (i.e., "wall," "roof," "door," "window"). The words were superimposed on three types of objects: faces with a neutral expression (same set of posers as used in the emotional task), houses, and cars. Sample stimuli are shown in Figure 2.

Behavioral Color–Word Stroop Task

The goal of this task (described in more detail in Friedman et al., 2016) was to index conflict between the prepotent tendency to read color words and the instruction to name

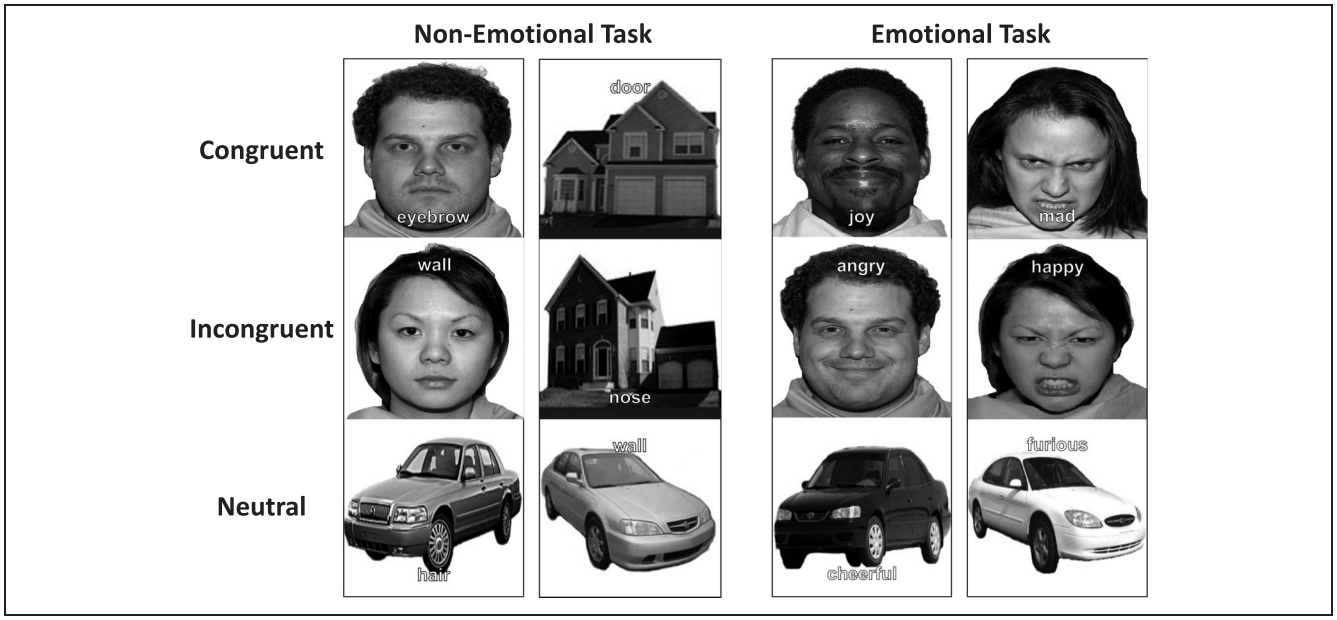

Figure 2. Examples of stimuli used in the non-emotional and emotional tasks. On each trial individuals viewed an item with a word superimposed on top of a picture and were required to make a category judgement (non-emotional: face vs. house, emotional: positive vs. negative) on the basis of the word.

the color in which the word was printed. On each trial, a colored stimulus (red, green, or blue) was presented on a black background and remained on the screen until the participant reported the color of the stimulus using the voice-activated E-prime microphone, followed by a 1000-msec intertrial interval. Participants first completed a block of 44 neutral trials (asterisks printed in colored ink) after 10 practice trials. Then, to engage the prepotency of word reading, participants completed a block of 44 congruent (word and color the same, e.g., "blue" in blue ink) trials, again preceded by 10 practice trials. Participants then completed a block of 62 incongruent trials (word and color do not match, e.g., "blue" in red ink) with no practice trials to minimize practice effects. Participants then completed three mixed blocks, which are the focus of the current analysis. These blocks consisted of all three trial types intermixed, with 56 trials per block, for a total of 168 trials (56 per condition).

For each subject and each trial type (e.g., mixed block incongruent trials), outlying RTs were identified by the Wilcox–Keselman trimming procedure (Wilcox & Keselman, 2003) and were excluded before calculating means. Stroop interference was calculated as the mean RT difference between the mixed block incongruent trials and the mixed block neutral trials.

Control Variables

Because of the broad age range among participants, age was included as a control variable. For participants with viable imaging data at both TP1 and TP2, age was averaged across the two time points. Age at a single time point was used for participants with only one time point of imaging data.

To minimize the possibility that any observed relationships between brain activation and Stroop interference could be driven by a participant's underlying processing speed, all models additionally included a measure of choice RT. Because the Stroop interference score is indexed by an RT difference score that captures the RT increase associated with more control-demanding trials, baseline processing speed should generally be removed from the score. However, in some cases, difference scores may still be impacted by their component RTs (Chapman, Chapman, Curran, & Miller, 1994). The choice RT task (Snyder & Hankin, 2016) required participants to press buttons with their right and left hands as fast as possible when presented with left or right pointing triangles for 60 trials. The dependent variable is mean RT.

Neuroimaging

Image Acquisition

Data were acquired on a Siemens MAGNETOM PRISMA 3.0 Tesla scanner with a 32-channel head coil at the Intermountain Neuroimaging Consortium on the campus of the University of Colorado Boulder for all participants except for 12 at the first time point for whom data were acquired on the pre-upgrade version of the same magnet (TIM TRIO). To reduce head motion during MRI data acquisition, foam padding was placed around participants' heads.

Structural scans were acquired via a T1-weighted magnetization prepared rapid gradient echo sequence in 224 sagittal slices, with a TR = 2400 msec, echo time = 2.07 msec, flip angle = 8°, field of view = 256 mm, and voxel size of .8 mm^3.

Functional scans were acquired via multiband accelerated T2*-weighted EPI (TR = 460 msec, echo time = 27.20 msec, flip angle = 44°, 56 interleaved axial slices aligned along anterior–posterior commissure line, slice thickness = 3 mm, field of view = 248 mm, multiband accel. Factor = 8, gap = 1 mm, voxel size = 3.0 mm^3). Two runs were acquired for each task at each time point, with each run consisting of 624 echo planar images, for 1248 images per task.

Preprocessing

fMRI preprocessing and GLM analyses were carried out using the FSL suite (Smith et al., 2004). For the Stroop tasks, the first 10 EPI volumes of each run were discarded to allow the MRI scanner to reach steady-state stability. Preprocessing included motion correction via ICA-based Automatic Removal Of Motion Artifacts (ICA-AROMA) (Pruim et al., 2015), an independent component analysis method for removing motion, high-pass filtering (100 sec), and brain extraction via Brain Extraction Tool (BET) (Smith, 2002). Registration of EPI images into subject- and standard-spaces was executed using FMRIB's Linear Image Registration Tool (FLIRT) (Jenkinson, Bannister, Brady, & Smith, 2002; Jenkinson & Smith, 2001). Individual subject EPI images were registered to that subject's magnetization prepared rapid gradient echo structural image via linear Boundary-Based Registration (Greve & Fischl, 2009) and then registered to the MNI-152 template via 12 degrees of freedom linear transformation. The resulting EPI images were smoothed using an 8-mm FWHM Gaussian smoothing kernel.

GLM Analyses

As blocked analyses can have up to 35% more power than event-related analyses (Bandettini & Cox, 2000), blocked contrasts were used in the GLM analyses. We have found in our prior studies that blocked contrasts in variants of the Stroop task enable individual differences to be detected (e.g., Andrews-Hanna et al., 2011; Banich et al., 2009). Because both the emotional and non-emotional tasks are identical in design, the same models were applied to both tasks.

FMRI Expert Analysis Tool (FEAT) within FSL was used to model the data. First-level models were run separately for each run of each task, at each time point, for each participant. These models consisted of three explanatory

variables, one for each block type: incongruent (Inc), congruent (Con), and neutral (Neu). Explanatory variables modeled the time of the entire block (i.e., also including neutral frequent trials). Fixation blocks (Fix) served as the nonmodeled baseline. The following contrasts of interest were created: Inc + Con + Neu > Fix, Inc + Con > Neu, Inc > Con, Inc > Neu. In cases where multiple trial types were combined within a contrast, average activation was computed across the trial types. Fixed effects analysis was used to estimate means across Run 1 and Run 2 of each task. Finally, four higher-level analyses were conducted to examine group main effects for both tasks at both time points.

Selection of ROIs

Bilateral brain ROIs were defined by parcels from within the Schaefer 400-area network parcellation (Schaefer et al., 2018), which reflects neurobiologically meaningful areas determined by resting-state fMRI across nearly 1500 participants. The 400-area parcellation was chosen as it provides an intermediate level of resolution that allows the areas of interest (e.g., IFJ) to be relatively isolated to a given parcel.

For each of the three ROIs, the right and left hemisphere parcels were determined according to the following steps. First, a peak voxel for each ROI was identified in the right hemisphere by the meta-analysis feature in Neurosynth (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). The term "stroop" was used to identify regions of IFJ that we posit maintain abstract task set representations and mid-cingulate regions that we posit carry out response selection in the face of conflicting information. To capture the mid-DLPFC regions that we posit maintain the working memory representation of specific items, the term "working memory" was used. In all cases, the Association Test maps, which identify those regions that are preferentially related to the search term, were used. One peak voxel for each ROI was determined using FSL's cluster function, and the Schaefer parcel containing the peak voxel was chosen to represent that ROI. The following right hemisphere peak voxels (reported in Montreal Neurological Institute coordinates) and parcels were identified: IFJ: 43, 12, 35, parcel 350; mid-cingulate: 7, 21, 25, parcel 311; mid-DLPFC: 42, 34, 30, parcel 351. The following three corresponding left hemisphere parcels were then visually identified: IFJ: parcel 141; mid-cingulate: parcel 107; mid-DLPFC: parcel 142.

Calculation of Beta Weights

Beta-weights for percent signal change were used to index brain activation. Beta-weights were calculated per contrast and participant within each of the six parcels using FSL's means function. Bilateral indices of activation were then calculated by taking the mean of the scaled right and left hemisphere parcel beta-weights for each of the three ROIs.

Approach and Analyses

Reliability of Brain Activation and Color–Word Stroop Behavioral Interference

We first investigated the reliability of whole-brain activation across all 400 parcels to better understand patterns of reliability across time points, tasks, and contrasts. Of particular interest are the following four measures: TP1 reliability across tasks, TP2 reliability across tasks, nonemotional task reliability across time points, and emotional task reliability across time points. Reliability was indexed by averaging across Fisher-transformed within-subject correlations. For example, TP1 reliability was computed by first calculating within-subject Pearson correlations between activation in the 400 parcels during the non-emotional task at TP1 with activation during the emotional task at TP1, Fisher-transforming each correlation value, averaging across subject-level correlations, and, finally, reverse transforming the average correlation value. The resulting correlation value can be interpreted as a standard correlation coefficient indexing the strength of the linear association between the two variables. Each reliability measure was calculated separately for Inc + Con + Neu > Fix, Inc + Con > Neu, and Inc > Con, resulting in 12 correlations (two time points × two tasks × three contrasts).

The same reliability patterns were assessed within the three ROIs. For this set of analyses, only a single Pearson correlation was computed for each reliability measure because each participant had only one activation value per parcel per contrast. Rather than computing reliability across all contrasts for all ROIs, we constrained these analyses to our a priori ROI by contrast pairings.

Because the behavioral Stroop task was performed only once at each time point, reliability was assessed by a single correlation between TP1 and TP2.

Conceptual Models

To evaluate the predicted relationships between individual differences in composite measures of brain activation and behavioral interference, we ran a single regression model containing each composite measure of brain activation as well as the two control variables (age and choice RT) as predictors. We additionally explored whether the same brain–behavior relationships were present when considering only a single observation, that is, one task at one time point. To do so, we ran four additional models in which the composite measures were replaced with measures calculated within each of the individual task and time points. All regressions were performed in R Version 4.2.2 (R Core Team, 2022).

Composite Measures of Brain Activation

To obtain more stable estimates of individual differences in brain activation, signal change in each parcel was

averaged across the emotional and non-emotional Stroop tasks at both TP1 and TP2. Because participants differed in their number of valid observation points, averages were computed from between one and four observation points. Participants with only one valid observation point ($n = 5$) were excluded from analyses. The same method was applied to the behavioral Stroop interference measure as well as control variables, with averages computed across TP1 and TP2. Participants with valid data at either TP1 or TP2 were included in analyses.

Analysis Model

The regression equation used to test the cascade-of-control model is presented in Equation (1).

$$\begin{aligned} Y = {} & \beta_0 + \beta_1(IFJ_{Inc+Con+Neu>Fix}) \\ & + \beta_2(Mid - DLPFC_{Inc+Con+Neu}) \\ & + \beta_3(Mid - Cingulate_{Inc>Con}) \\ & + \beta_4(IFJ_{Inc+Con+Neu>Fix} * Mid - Cingulate_{Inc>Con}) \\ & + \beta_5(ChoiceRT) + \beta_6(Age) + \varepsilon \end{aligned} \tag{1}$$

All variables were standardized such that the mean is 0 and the standard deviation was 1. Y represents the outcome measure, Stroop interference. β_1 captures the expected increase in Stroop interference associated with a 1 *SD* increase in activation of IFJ for an individual with average mid-cingulate activation. β_2 captures the interference increase associated with 1 *SD* increases of mid-DLPFC activation. β_3 captures the expected interference increase associated with a 1 *SD* increase in activation of mid-cingulate for an individual with average IFJ activation. β_4 captures the interaction between IFJ and mid-cingulate activation, which is a prominent feature of the cascade-of-control model, which posits that cingulate activation is influenced by the degree of top–down biasing by IFJ (e.g., Silton et al., 2010). β_5 and β_6 capture the two control variables, age and choice RT, respectively.

Bootstrapped Confidence Intervals

To obtain a quantitative measurement of the uncertainty around each beta estimated by the model above, we constructed bootstrapped confidence intervals. Specifically, we ran the model across 10,000 iterations in which we randomly sampled, with replacement, 62 data points from our data set ($n = 62$). We then used the set of beta estimates extracted across the 10,000 iterations to generate a sampling distribution and confidence interval for each effect.

Alternative ROI by Contrast Pairings

To test the specificity of the hypothesized ROI by contrast pairings, we compared the cascade-of-control model to models representing all alternative contrast pairings. To test each pairing, we retained the model structure with regard to the ROIs but assigned each of the ROIs one of the following contrasts: Inc + Con + Neu > Fix, Inc + Con > Neu, Inc > Neu, Inc > Con. This process resulted in 64 possible contrast pairings (4 IFJ pairings by 4 mid-DLPFC pairings by 4 mid-cingulate pairings).

We compared the goodness of fit of the cascade-of-control model against each alternative model by computing the difference in Bayesian information criterion (ΔBIC). ΔBIC is a commonly used criterion for selecting among models and is the preferred metric when explanatory goodness of fit is desired over predictive accuracy (Sober, 2002). ΔBIC was calculated by subtracting the cascade-of-control model BIC from the alternate model BIC, such that positive numbers indicate better fit for the cascade model. Based on guidelines set by Raftery (1995), we consider a difference greater than 10 to be strong evidence of a meaningful difference.

RESULTS

Behavioral Results

To ensure that each of the tasks engendered cognitive control as intended, we assessed patterns of behavioral performance within each task at each time point. RT difference scores and accuracy data for the fMRI tasks and the Color–Word Stroop task are presented in Table 1. Consistent with expected task performance, one-sample t tests indicated significant RT interference for each task at each time point (ps < .01).

Reliability

Whole Brain

To better understand the reliability of brain activation across time points, tasks, and contrasts within our data, we first investigated whole-brain reliability. The general pattern of results is described here, and full results are presented in Table 2. Reliability differed substantially across contrasts, with minimal differences observed in relation to time point and task. Reliability was highest (.84–.89) for the Inc + Con + Neu > Fix contrast, followed by the Inc + Con > Neu contrast (.27–.32) and the Inc > Neu contrast (.24–.33). Reliability was poor for the Inc > Con contrast for both the time point and task measures (−.01 to .15); however, within-task reliability for both the non-emotional task (.11) and emotional task (.15) was numerically higher than the within-time point reliability at both TP1 (.04) and TP2 (−.01).

Within-ROI

Within-ROI reliability largely followed the same contrast-dependent pattern seen in the whole-brain results (see Table 2). Reliability was highest within the IFJ ROI for the Inc + Con + Neu > Fix contrast with significant correlations observed across all task and time point reliability

Table 1. Behavioral Results

		RT Interference		*Inc Accuracy*	
Time Point	*Task*	*M*	*SD*	*M*	*SD*
TP1	Non-emotional (Inc-Con)	55.56	32.09	93.89	5.09
	Emotional (Inc-Con)	39.83	31.77	94.91	4.24
	Color–Word (Inc-Neu)	158.46	91.39	97.37	2.92
TP2	Non-emotional (Inc-Con)	48.30	25.91	94.95	3.62
	Emotion (Inc-Con)	44.89	29.93	93.57	5.15
	Color–Word (Inc-Neu)	147.68	87.85	97.94	2.58

RT is reported in milliseconds, and accuracy is reported as a percentage.

measures (.50–.80, ps < .001). For this contrast, within-time point reliability appeared to be somewhat higher (.65 and .80) compared with within-task reliability (.50 and .70). Reliability within the mid-DLPFC ROI for the Inc + Con > Neu contrast varied considerably (−.18 to .47), but there was no clear pattern with regard to time point or task. Mid-cingulate reliability for the Inc > Con contrast was poor (−.13–.16).

Color–Word

The Color–Word Stroop I-N RT interference score showed moderate reliability from TP1 to TP2, $r(48) = .56, p < .001$.

Composite Measures Model

We first examined how well an individual's behavioral Stroop interference could be predicted by a model that included composite activation for the contrast of interest for each of the three ROIs as well as an interaction term that examined whether activation in IFJ moderated the relationship between mid-cingulate activation and interference. Model results are presented in Table 3. The full model explained a significant amount of variance in interference scores, $F(6, 55) = 3.15\ p = .010$, $R^2 = .26$, $R^2_{\text{Adjusted}} = .17$.

Mean activation in IFJ (Inc + Con + Neu > Fix) was significantly positively associated with interference such that greater activation was associated with more interference ($\beta = 0.31, t(55) = 2.43, p = .018$). Mid-DLPFC (Inc + Con > Neu) was also significantly positively associated with interference ($\beta = 0.34, t(55) = 2.73, p = .009$). As predicted, a significant interaction was observed such that mid-cingulate activation positively predicted behavioral interference when IFJ activation was low ($\beta = -0.36, t(55) = -2.83, p = .007$). Analysis of simple slopes

Table 2. Reliability

		Within-ROI Correlations			*Whole-Brain Average Correlations*				
	n	*IFJ C + I + N > Fix*	*Mid-DLPFC C + I > N*	*Mid-Cingulate I > C*	*C + I + N > Fix*	*C + I + N > Fix*	*C + I > N*	*I > N*	*I > C*
Non-emotional Task × Emotional Task at TP1	53	.65***	−.02	−.13		.89	.27	.28	.04
Non-emotional Task × Emotional Task at TP2	38	.80***	.47***	−.12		.87	.27	.24	−.01
TP1 × TP2 in Non-emotional Task	34	.70***	.35*	.16		.88	.30	.28	.11
TP1 × TP2 in Emotional Task	41	.50***	−.18	.07		.84	.32	.33	.15

Significance levels are not provided for whole-brain reliability measures as they reflect average correlations.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 3. Cascade-of-Control Model Results

Predictor	*Beta*	*SE*	*t-statistic*	*95% CI*	*p*
Intercept	0.02	0.12	0.14	−0.22, 0.25	.890
IFJ C + I + N > Fix	0.31	0.13	2.43	0.06, 0.57	.018*
Mid-DLPFC C + I > N	0.34	0.13	2.73	0.09, 0.60	.009**
Mid-cingulate I > C	−0.02	0.12	−0.17	−0.26, 0.22	.863
IFJ × Mid-cingulate interaction	−0.36	0.13	−2.83	−0.61, −0.10	.007**
Choice RT	−0.09	0.12	−0.69	−0.33, 0.16	.495
Age	0.27	0.13	2.17	0.02, 0.52	.035*

* $p < .05$.

** $p < .01$.

*** $p < .001$.

revealed a significant positive association between mid-cingulate activation and behavioral interference at low levels of IFJ activation (−1 *SD*; β = 0.33, $t(55) = 2.03$, $p = .047$), and a negative association at high levels of IFJ activation (+1 *SD*; β = −0.38, $t(55) = -2.10$, $p = .040$). Age positively predicted interference such that interference tended to be higher in older participants (β = 0.27, $t(55) = 2.17$, $p = .035$). Choice RT was not associated with interference ($p = .495$).

Bootstrapped Confidence Intervals

Bootstrapping supported the findings reported in the previous section. The 95% confidence intervals for IFJ activation (0.03–0.61), mid-DLPFC activation (0.13–0.59), and the interaction between IFJ and mid-cingulate activation (−0.61 to −0.14) did not contain zero. Similarly, the 95% confidence interval for age also did not contain zero (0.03–0.54). See Figure 4.

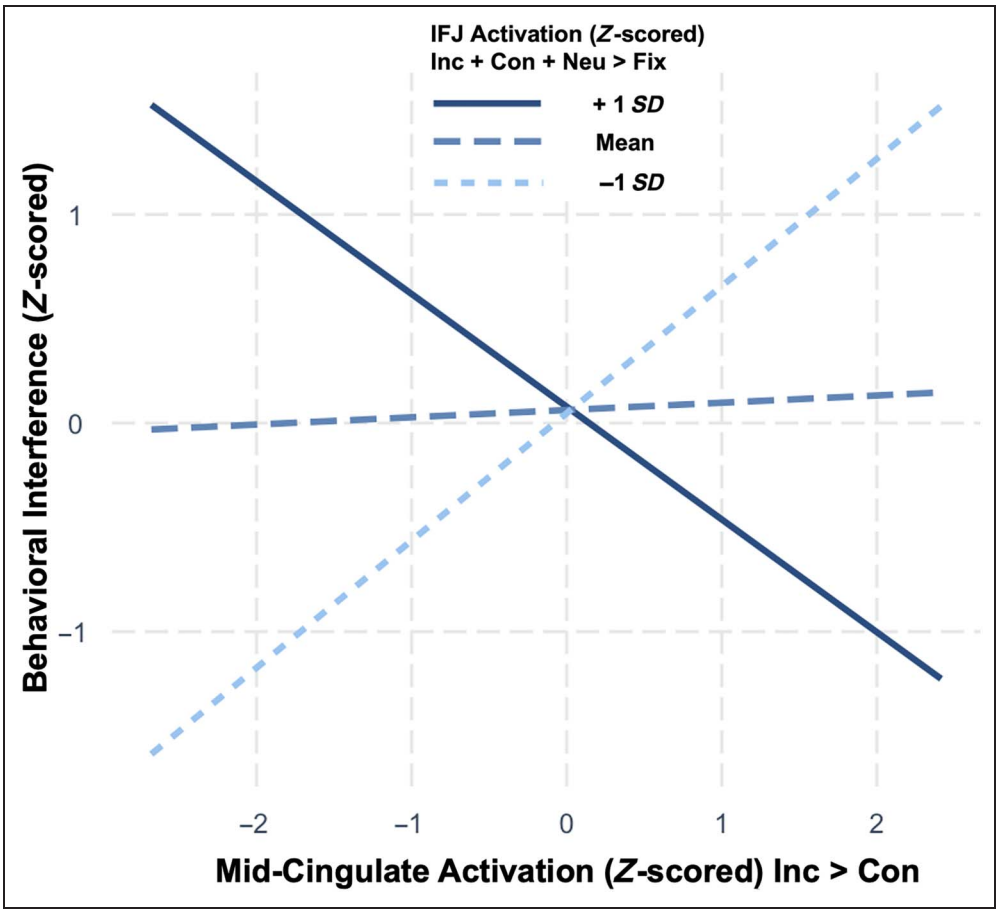

Figure 3. The relationship between mid-cingulate activation and behavioral interference as a function of IFJ activation. Plot created using *interact_plot* from package interactions Version 1.1.5 (Long, 2022).

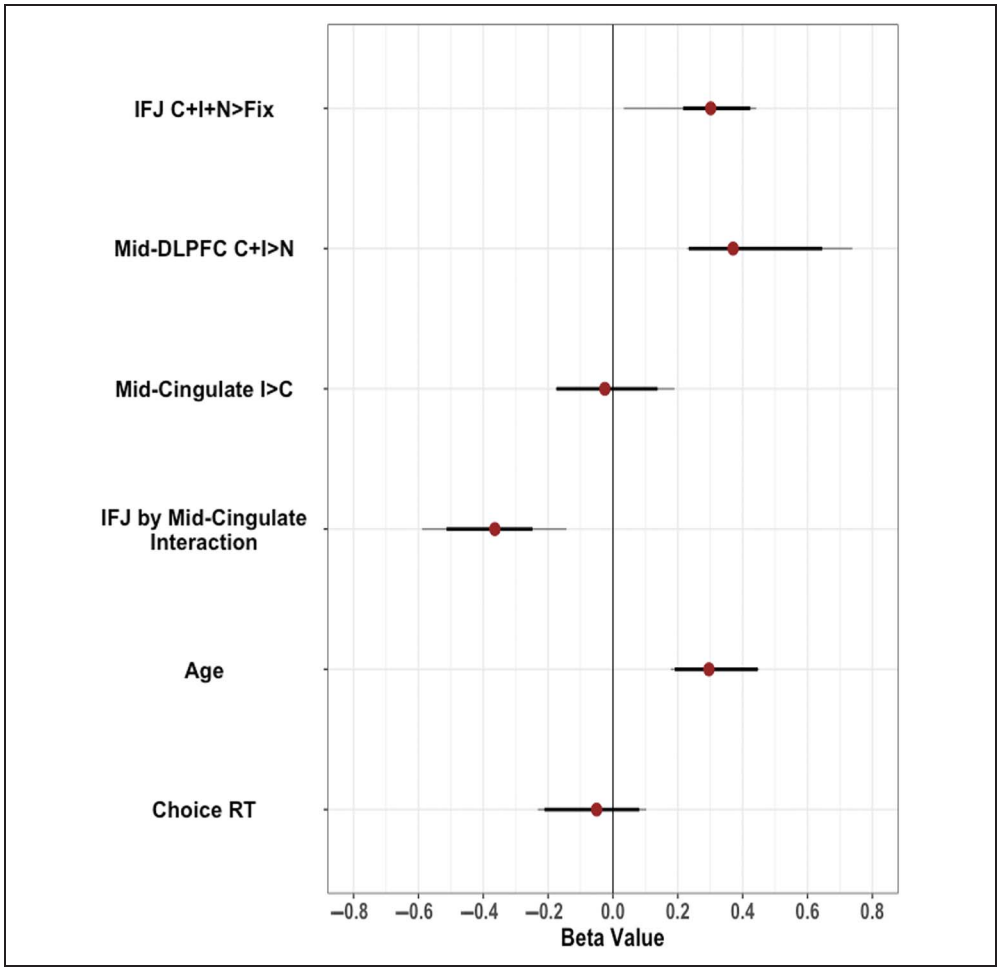

Figure 4. Bootstrap distributions of beta estimates from the cascade-of-control model. The predictor variables from the linear regression are shown along the *y* axis. Red circles indicate the mean beta estimate. Black lines and gray lines indicate the 80% and 95% confidence intervals, respectively.

Alternative ROI by Contrast Pairings

We next evaluated the performance of the cascade-of-control model against 63 models with alternative contrast pairings. The models with the 10 lowest BICs are reported in Table 4. The contrast pairings hypothesized by the cascade-of-control model resulted in the lowest BIC of all 64 combinations (BIC = 189.67). The model with the

Table 4. Model Comparisons

ROI by Contrast Pairing				
IFJ	*Mid-DLPFC*	*Mid-cingulate*	*BIC*	*ΔBIC*
C + I + N > Fix	C + I > N	I > C	189.67	–
C + I + N > Fix	I > N	I > C	190.89	1.22
C + I + N > Fix	I > N	C + I > N	194.04	4.37
C + I + N > Fix	C + I > N	C + I > N	194.21	4.54
C + I + N > Fix	I > C	C + I > N	194.50	4.83
C + I + N > Fix	C + I + N > Fix	C + I > N	194.65	4.98
C + I + N > Fix	C + I > N	I > N	196.45	6.78
C + I + N > Fix	C + I + N > Fix	I > C	196.60	6.93
C + I + N > Fix	I > N	I > N	196.71	7.04
C + I + N > Fix	C + I + N > Fix	I > N	196.87	7.20

ROI by contrast pairings resulting in the 10 best goodness of fit metrics. ΔBIC was calculated by subtracting the lowest BIC from that of each alternative model.

Table 5. Single Observation Point Model

Time Point	*Task*	*n*	*Predictor*	*Beta*	*SE*	*t-statistic*	*95% CI*	*p*
TP1	Non-emotional	53	Intercept	0.02	0.13	0.11	−0.25, 0.28	.911
			IFJ C + I + N > Fix	0.25	0.14	1.72	−0.04, 0.53	.093
			Mid-DLPFC C + I > N	0.29	0.14	2.13	0.02, 0.56	.039*
			Mid-cingulate I > C	−0.03	0.14	−0.20	−0.32, 0.26	.845
			IFJ C + I + N > Fix × Mid-cingulate I > C	−0.12	0.16	−0.77	−0.45, 0.20	.448
			Choice RT	−0.04	0.14	−0.30	−0.32, 0.24	.765
			Age	0.20	0.14	1.38	−0.09, 0.48	.176
	Emotional	55	Intercept	0.01	0.13	0.09	−0.24, 0.27	.931
			IFJ C + I + N > Fix	0.28	0.14	2.02	0.00, 0.57	.049*
			Mid-DLPFC C + I > N	0.10	0.14	0.74	−0.17, 0.38	.460
			Mid-cingulate I > C	−0.25	0.13	−1.92	−0.52, 0.01	.060
			IFJ C + I + N > Fix × Mid-cingulate I > C	−0.19	0.13	−1.51	−0.45, 0.06	.137
			Choice RT	0.06	0.14	0.41	−0.23, 0.35	.680
			Age	0.25	0.14	1.81	−0.03, 0.54	.076
TP2	Non-emotional	40	Intercept	0.00	0.17	0.00	−0.34, 0.34	.998
			IFJ C + I + N > Fix	0.17	0.19	0.88	−0.22, 0.55	.387
			Mid-DLPFC C + I > N	0.03	0.19	0.17	−0.35, 0.41	.867
			Mid-cingulate I > C	−0.08	0.19	−0.44	−0.47, 0.30	.664
			IFJ C + I + N > Fix × Mid-cingulate I > C	−0.16	0.16	−1.00	−0.50, 0.17	.326
			Choice RT	−0.09	0.18	−0.47	−0.46, 0.28	.639
			Age	0.17	0.19	0.94	−0.20, 0.55	.355
	Emotional	45	Intercept	−0.03	0.15	−0.19	−0.34, 0.28	.854
			IFJ C + I + N > Fix	0.18	0.16	1.14	−0.14, 0.51	.262
			Mid-DLPFC C + I > N	0.15	0.16	0.94	−0.17, 0.46	.355
			Mid-Cingulate I > C	0.13	0.17	0.79	−0.21, 0.48	.433
			IFJ C + I + N > Fix × Mid-cingulate I > C	0.22	0.19	1.15	−0.17, 0.61	.257
			Choice RT	−0.03	0.16	−0.19	−0.36, 0.30	.853
			Age	0.07	0.17	0.43	−0.26, 0.41	.668

* $p < .05$.

** $p < .01$.

*** $p < .001$.

second lowest BIC was identical except that activation within mid-DLPFC was indexed by Inc > Neu, rather than Inc + Con > Neu (BIC = 190.89). Comparing the cascade model to the third alternative model resulted in a ΔBIC of 4.37, demonstrating moderately better fit for the cascade model.

Single Observation Models

Finally, we assessed whether the same relationships were observed when data from only a single task and single time point were used to index brain activation. No consistent brain–behavior relationships were observed across the four single observation models. Significant results are reported in the text, and full results for each model are presented in Table 5. For the non-emotional task at TP1, mid-DLPFC activation showed a significant positive relationship with behavioral interference ($\beta = 0.29$, $t(46) = 2.13$, $p = .039$). For the emotional task within the same session, a significant relationship was observed between IFJ activation and behavioral interference ($\beta = 0.28$, $t(48) = 2.02$, $p = .049$). No significant relationships were observed for either task at TP2 ($ps > .1$).

DISCUSSION

Overview

The results of the present study are largely consistent with the cascade-of-control model (Banich, 2019), providing evidence that cognitive control is enacted via multiple regions of frontal cortex that each resolve specific levels of interference in a cascade-like manner. We demonstrated that individual differences in composite measures of brain activation within three ROIs were associated with behavioral interference scores on a separate, out-of-scanner task. The present study is novel in that we not only confirmed the interactive brain–behavior relationships hypothesized by the cascade-of-control model but also showed that these relationships are linked to brain activation indexed at specific levels of interference (for other evidence that distinct brain regions may code various aspects of stimulus properties and control processes related to performance on the Stroop task, see Freund, Bugg, & Braver, 2021). Furthermore, this is one of relatively few studies to use composite measures of brain activation collected across multiple fMRI tasks and multiple time points. Below, we discuss the current findings as they relate to the cascade-of-control model as well as the variation between the brain–behavior relationships observed in the composite measures model as compared with the single observation models, and the potential implications of these differences.

Brain–Behavior Relationships

IFJ

Consistent with the cascade-of-control model, individuals with less activation in IFJ showed significantly lower levels of behavioral interference on the out-of-scanner Stroop task. Although the directionality of activation differences can be somewhat convoluted (e.g., see Rypma et al., 2006), we interpret lower levels of activation in this region as evidence of more efficient task set maintenance.

Also as predicted, the relationship between IFJ and behavioral interference depended on mid-cingulate activation. Based on prior work demonstrating that task sets are maintained regardless of trial type (e.g., Milham et al., 2001), we additionally predicted and confirmed that this relationship is specific to the Inc + Con + Neu > Fix contrast pairing.

Mid-DLPFC

A significant main effect was observed within mid-DLPFC such that individuals with less activation in mid-DLFPC showed significantly lower levels of behavioral interference. We interpret lower levels of activation in mid-DLPFC as evidence of more efficient selection of information in working memory. On the other hand, for those individuals who have difficulty resolving competition, biasing within mid-DLPFC must be ramped up to successfully complete the task, and thus activation is greater.

Also in line with our predictions, this mid-DLPFC relationship was specific to the Inc + Con > Neu contrast. We had proposed that mid-DLPFC effects would be linked to this contrast because for both incongruent and congruent trials, response-relevant information is present in both the task-relevant and task-irrelevant stimulus dimensions (Milham & Banich, 2005). As such, both trial types should theoretically require the selection of two pieces of information that are being maintained in working memory. This contrasts with neutral trials in which there is only one source of response-relevant information that is contained in the ink color. The current findings are also consistent with a body of prior work demonstrating that regions of mid-DLPFC may reduce interference from competing information by buffering relevant information in working memory (e.g., Burgess & Braver, 2010).

Mid-cingulate

As hypothesized, the relationship between mid-cingulate activation and behavioral interference was moderated by IFJ activation. Results further supported the hypothesized Inc > Con contrast pairing, which captures the late-stage response selection carried out by mid-cingulate.

Interaction between IFJ and Mid-cingulate

Central to the cascade-of-control model is the assertion that successful competition resolution becomes dependent on late-stage response selection in mid-cingulate regions when a task set is not adequately maintained and implemented within IFJ. Consistent with this tenet, a significant interaction was observed between mean

activation in IFJ (Inc + Con + Neu > Fix) and mean activation in mid-cingulate (Inc > Con). As expected, behavioral interference was highest for individuals with high levels of IFJ activation coupled with low levels of mid-cingulate activation. As levels of mid-cingulate activation increased, the resulting level of behavioral interference decreased. We interpret high levels of IFJ activation as evidence of more effortful and poorer task set implementation, which likely results in proportionally less task-relevant information being fed downstream. Without adequate biasing having occurred previously within the cascade, the resolution of interference must occur at the final stage of the cascade, that is, as late-stage response selection within mid-cingulate. For those individuals who do not successfully ramp up mid-cingulate to "pick up the slack," interference remains high.

The full crossover nature of the interaction was not predicted and is somewhat intriguing. As expected, individuals who efficiently implemented a task set within IFJ showed low behavioral interference with minimal engagement of mid-cingulate. However, our results indicate that ramping up mid-cingulate is not just unnecessary in this case, but that doing so may actually be detrimental when efficient biasing has already occurred within IFJ. One potential interpretation of this pattern is that these individuals are employing more caution at the final stage at which control can be implemented. That is, control is being exerted over response selection, even when the correct response is sufficiently activated. If this were the case, then we might expect a speed-accuracy trade-off in which increases in mid-cingulate activation were accompanied by increases in accuracy. Such speed-accuracy trade-offs are commonly observed in cognitive control experiments and have been previously reported in work similar to the current study (Silton et al., 2010). Unfortunately, error rates in the current study were insufficient for analysis and so this remains an open question for future investigation.

Although the current study was not designed to interrogate the temporal directionality of the relationship between IFJ and mid-cingulate activation, the observed interaction is consistent with prior work supporting a directional path of information from lateral to medial prefrontal regions. Using a combined fMRI and ERP methodology, Silton et al. (2010) also observed a significant interaction between activity within posterior regions of dorsolateral pFC (the same area referred to as IFJ in the current study) and cingulate regions. Critically, source-waveform ERP mediation analysis revealed that the relationship between activity within left posterior dorsolateral pFC and behavioral interference was mediated by activity in right dorsal ACC. As hypothesized by the cascade-of-control model, this pattern was specific to early activity within posterior DLPFC (300–400 msec post-presentation) and late activity within anterior cingulate (520–680 msec post-presentation), with alternative models of temporal information flow (e.g., early ACC and late posterior DLPFC) not supported. Thus, we tentatively interpret the interaction between IFJ and mid-cingulate activation in the current study as further evidence that successful implementation of a task set within IFJ modulated the need to resolve response-related interference in mid-cingulate.

Alternative ROI by Contrast Pairings

When tested against all possible alternative ROI by contrast pairings, the cascade-of-control model demonstrated better fit than each of the 63 alternative models. Perhaps most interesting is the cascade model's outperformance of an alternative model in which activation across all ROIs was indexed by the Inc > Neu contrast. Because behavioral interference was indexed by the RT difference between incongruent and neutral trials, it would be reasonable to assume that the same difference in ROI activation, that is, the Inc > Neu contrast, would yield the strongest associations. However, this does not appear to be true in the current data. Similarly, an argument could be made for using the Inc > Fix contrast to capture overall neural engagement during the most effortful trials or, alternatively, the Inc > Con contrast to capture activation that occurs specifically when conflicting information is present in both the task-relevant and task-irrelevant dimensions. However, these models are again not supported by the data. Taken together, these results suggest that the best index of activation for a particular region is that which captures the cognitive operation(s) most likely to be dependent on that region.

Composite Versus Single Observation Measures of Brain Activation

Whereas the results of the composite measures model were consistent with the cascade-of-control model, similar results were not observed in the single observation models. The results of these four models were neither consistent with our theoretical model nor with one another. Notably, only two significant brain–behavior relationships were present in the single observation models. At TP1, a positive relationship was observed between mid-DLPFC and behavioral interference for the non-emotional task and a significant relationship between IFJ and interference was observed for the emotional task.

As discussed previously, reliability across task and time point varied considerably. Although there was some degree of consistency, particularly for the task versus fixation contrasts, it appears that the single time point and single task measures of activation contain a substantial amount of noise. There are numerous potential sources of noise ranging from random measurement error in both the brain and behavior measures to state-specific factors such as alertness, which may have varied across testing sessions. Although it is difficult to identify a specific source, averaging across measures may have resulted in

a level of noise reduction sufficient for the observed brain–behavior relationships to emerge.

Although not the focus of the present study, the discrepancies between the composite model and the single observation models have potential implications, most notably for individual difference studies. Although the growing trend has been to increase sample sizes (e.g., ABCD, Casey et al., 2018), there may also be considerable value in increasing the quantity of data collected from each individual participant. This deep phenotyping approach has been used across a variety of cognitive neuroimaging applications. For example, within a small sample of 10 participants, Gordon et al. (2017) collected 6 hr of fMRI data across three tasks in addition to multiple anatomical and resting-state scans. The resulting Midnight Scan Club data set has allowed for the mapping of neural organization at the individual participant level, as well as numerous other investigations of functional network characteristics (e.g., Dworetsky et al., 2021). A similar approach has been employed by others including King, Hernandez-Castillo, Poldrack, Ivry, and Diedrichsen (2019) who used a battery of 26 tasks to map functional boundaries of the cerebellum in 24 participants. The results of the current study similarly provide some support for this approach as an effective way to better identify meaningful brain–behavior relationships.

There are, however, downsides to collecting larger quantities of data per participant including more hours spent in the laboratory, potentially over multiple sessions, which can greatly increase the burden on participants. Furthermore, issues such as missing data because of attrition and poor data quality because of fatigue may arise. In addition, this approach may necessitate a smaller sample size, raising generalizability concerns. Ultimately, the best method for testing individual differences will depend on the experimental question and the available resources.

Reliability

The reliability results observed within the current study are consistent with prior work demonstrating that fMRI task contrasts that show robust within-subject differences often do not demonstrate high reliability (e.g., Kennedy et al., 2022; Infantolino, Luking, Sauder, Curtin, & Hajcak, 2018). Across both the whole-brain and ROI measures, reliability estimates differed substantially with the Inc + Con + Neu > Fix contrast showing the highest levels of reliability and the classical Inc > Con contrast showing very poor levels of reliability. These results are likely driven in part by the statistical nature of fMRI contrasts in which one task condition is subtracted from another task condition. Although critical from a theoretical perspective, the use of difference scores to isolate various cognitive processes of interest negatively impacts measurement reliability because the reliability of difference scores will always be lower than the reliability of each component measure (Hedge, Powell, & Sumner, 2018). As discussed by Infantolino et al. (2018), high correlations between component measures contribute substantially to the poor reliability of any type of difference score. We see evidence of this pattern in our results as reliability is lowest for those task contrasts in which the cognitive and mental processes required by each of the contrast components are most similar, that is, incongruent and congruent trials.

Despite well-established concerns about their reliability, the findings of the current study suggest that individual differences in condition versus condition task contrasts may reflect meaningful variation. Therefore, the use of strategies to increase reliability may be of particular importance for individual differences studies in which isolating specific cognitive processes is of particular interest. Multiple such strategies including the use of composite measures were employed in the current study; however, the suitability and value of any strategy will depend on many study factors (see Elliott et al., 2021).

Comparisons with Other Conceptualizations of Cognitive Control as Implemented by Prefrontal Regions

Although the cascade-of-control model was found in the current study to be useful in revealing brain–behavior relationships, there are other models of prefrontal regions involved in cognitive control that might also be considered and that differ in important ways from the model utilized here. For example, Dosenbach et al. (2006) argue that rather than IFJ imposing a task set, the regions responsible for sustaining task sets are the dorsal ACC/medial superior frontal cortex and bilateral anterior insula/frontal operculum. Determining which of these two models is more appropriate is beyond the scope of the current paper.

In contrast to our model which posits that IFJ serves as the critical first point of the cascade, models of cognitive control by Badre and Nee (2018) suggest that mid-DLPFC sits at the top of the hierarchy of control, although they do not make specific arguments about whether that means these regions acts before or after other regions. The model of Koechlin and colleagues (Azuar et al., 2014; Koechlin & Summerfield, 2007) argues for a somewhat different organization with levels of control organized in a rostral to caudal manner over pFC, with more posterior regions involved in sensory control, middle regions involved in contextual control, and more anterior regions in episodic control. Our model is consistent with aspects of this conceptualization as it assumes that control in IFJ precedes that of mid-DLPFC.

Our model also is distinct in important ways from the model of Botvinick and colleagues (Botvinick, Cohen, & Carter, 2004; Botvinick, Braver, Barch, Carter, & Cohen, 2001). In that model, the emphasis is on the anterior cingulate acting as a conflict detector, which, when activated, sends a signal to DLPFC to increase control. Our model is distinct in that we have argued that the role of the mid-

cingulate region is mainly to exert control at the level of response output, rather than being a conflict detector more generally (e.g., Milham & Banich, 2005; Milham et al., 2001). In addition, our emphasis is on the directional relationship in time between activity in mid-DLPFC and ACC rather than the other way around as emphasized in the Botvinich model (as an aside, the ACC to DLPFC feedback loop of Botvinick and colleagues could be easily integrated into the cascade-of-control model). Future work will be needed to adjudicate the degree to which these various models provide contradictory or complimentary perspectives that allow individual differences in brain activation to be linked to individual differences in behavioral performance.

Limitations

There are several limitations to the current study. First, our sample size was relatively small for an investigation of individual differences. Second, the inclusion of participants with as few as two observation points is suboptimal as many of these participants contributed data at only one time point or for only one task. However, because of the trade-off with sample size, we determined this to be the optimal cutoff. Third, our sample consisted only of women and, as such, it is possible the observed brain–behavior relationships may not generalize to men, yet the likelihood of substantial sex differences in the present findings is low as multiple studies have observed no such sex differences in performance on Stroop tasks (see Gaillard, Fehring, & Rossell, 2021). Fourth, the current study cannot provide definitive evidence for the cascade-of-control model, which argues for a directional temporal relationship between activation in different brain regions (i.e., IFJ activation influencing later processing in ACC). fMRI is ill-suited for examining such temporal relationships compared with other cognitive neuroscience methods such as electrophysiological recordings, which can more fruitfully address this issue (see, e.g., Silton et al., 2010). Nonetheless, the model predicts an interaction effect between activation in IFJ and ACC, which was observed in the current study. Fifth, although the Stroop task is considered a "gold standard" task of cognitive control (MacLeod, 1992), the degree to which our model is able to predict brain–behavior relationships in other cognitive control tasks remains to be seen.

Summary

The present study demonstrated that individual differences in behavioral Stroop interference measured outside the magnet are associated with composite measures of brain activation in a manner consistent with the cascade-of-control model. As hypothesized, activation within IFJ and mid-DLPFC regions predicted performance on an out-of-scanner Stroop task. Furthermore, a significant interaction between IFJ and mid-cingulate provided evidence that successful engagement of a task set within IFJ reduces the need for late-stage response selection within mid-cingulate regions. In addition, alternative models could not adequately predict brain–behavior relationships in comparison to the theoretically derived ROI by contrast pairings that reflect the cognitive processes hypothesized to occur within each region by the cascade-of-control model. These findings provide evidence that individual differences in engagement of multiple regions of frontal cortex, each of which contributes differently to resolving different types of interference, may help to explain individual differences in cognitive control outside the neuroimaging environment. Finally, the current study provides an example of how a "hybrid" approach to obtaining reliable and stable measures that considered both between- and within-subject variance through the use of multiple measures and data collection time points may allow for the detection of brain–behavior relationships.

Acknowledgments

We acknowledge the help of David Caha, Kenny Carlson, Rebecca Helmuth, Kathy Pearson, and Amy Hegarty, PhD for their help in data collection, organization, and preliminary aspects of analyses, as well as the staff of the Intermountain Neuroimaging Consortium, especially Nicole Speer, Operations Director, and Teryn Wilkes, Head MR Technologist.

Reprint requests should be sent to Louisa L. Smith, Department of Psychology and Neuroscience, D444 Muenzinger Hall, UCB 345, University of Colorado Boulder, Boulder, Colorado 80309, or via e-mail: Louisa.Smith@colorado.edu.

Data Availability Statement

Data and analysis materials are available upon request.

Author Contributions

Louisa L. Smith: Conceptualization; Formal analysis; Visualization; Writing—Original draft; Writing—Review & editing. Hannah R. Snyder: Funding acquisition; Methodology; Writing—Review & editing. Benjamin L. Hankin: Funding acquisition; Methodology; Writing—Review & editing. Marie T. Banich: Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing—Review & editing.

Funding Information

Marie T. Banich and Benjamin L. Hankin, National Institute of Mental Health (https://dx.doi.org/10.13039/100000025), grant number: R01MH105501.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent

pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

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*, e21598. https://doi.org/10.1371/journal.pone.0021598, PubMed: 21738725

Azuar, C., Reyes, P., Slachevsky, A., Volle, E., Kinkingnehun, S., Kouneiher, F., et al. (2014). Testing the model of caudo-rostral organization of cognitive control in the human with frontal lesions. *Neuroimage*, *84*, 1053–1060. https://doi.org/10.1016/j.neuroimage.2013.09.031, PubMed: 24064070

Badre, D., & Nee, D. E. (2018). Frontal cortex and the hierarchical control of behavior. *Trends in Cognitive Sciences*, *22*, 170–188. https://doi.org/10.1016/j.tics.2017.11.005, PubMed: 29229206

Baker, D. H., Vilidaite, G., Lygo, F. A., Smith, A. K., Flack, T. R., Gouws, A. D., et al. (2021). Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods*, *26*, 295–314. https://doi.org/10.1037/met0000337, PubMed: 32673043

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, *344*, 424–427. https://doi.org/10.1126/science.1247003, PubMed: 24763592

Bandettini, P. A., & Cox, R. W. (2000). Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magnetic Resonance in Medicine*, *43*, 540–548. https://doi.org/10.1002/(SICI)1522-2594(200004)43:4<540::AID-MRM8>3.0.CO;2-R, PubMed: 10748429

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science*, *18*, 89–94. https://doi.org/10.1111/j.1467-8721.2009.01615.x

Banich, M. T. (2019). The Stroop effect occurs at multiple points along a cascade of control: Evidence from cognitive neuroscience approaches. *Frontiers in Psychology*, *10*, 2164. https://doi.org/10.3389/fpsyg.2019.02164, PubMed: 31681058

Banich, M. T., Mackiewicz, K. L., Depue, B. E., Whitmer, A. J., Miller, G. A., & Heller, W. (2009). Cognitive control mechanisms, emotion and memory: A neural perspective with implications for psychopathology. *Neuroscience & Biobehavioral Reviews*, *33*, 613–630. https://doi.org/10.1016/j.neubiorev.2008.09.010, PubMed: 18948135

Banich, M. T., Milham, M. P., Atchley, R. A., Cohen, N. J., Webb, A., Wszalek, T., et al. (2000). Prefrontal regions play a predominant role in imposing an attentional 'set': Evidence from fMRI. *Cognitive Brain Research*, *10*, 1–9. https://doi.org/10.1016/S0926-6410(00)00015-X, PubMed: 10978687

Banich, M. T., Smolker, H. R., Snyder, H. R., Lewis-Peacock, J. A., Godinez, D. A., Wager, T. D., et al. (2019). Turning down the heat: Neural mechanisms of cognitive control for inhibiting task-irrelevant emotional information during adolescence. *Neuropsychologia*, *125*, 93–108. https://doi.org/10.1016/j.neuropsychologia.2018.12.006, PubMed: 30615898

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652. https://doi.org/10.1037/0033-295X.108.3.624, PubMed: 11488380

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539–546. https://doi.org/10.1016/j.tics.2004.10.003, PubMed: 15556023

Brass, M., Derrfuss, J., Forstmann, B., & von Cramon, D. Y. (2005). The role of the inferior frontal junction area in cognitive control. *Trends in Cognitive Sciences*, *9*, 314–316. https://doi.org/10.1016/j.tics.2005.05.001, PubMed: 15927520

Burgess, G. C., & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: Effects of interference expectancy and fluid intelligence. *PloS One*, *5*, e12861. https://doi.org/10.1371/journal.pone.0012861, PubMed: 20877464

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., et al. (2018). The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001, PubMed: 29567376

Chapman, L. J., Chapman, J. P., Curran, T. E., & Miller, M. B. (1994). Do children and the elderly show heightened semantic priming? How to answer the question. *Developmental Review*, *14*, 159–185. https://doi.org/10.1006/drev.1994.1007

Chechlacz, M., Gillebert, C. R., Vangkilde, S. A., Petersen, A., & Humphreys, G. W. (2015). Structural variability within frontoparietal networks and individual differences in attentional functions: An approach using the theory of visual attention. *Journal of Neuroscience*, *35*, 10647–10658. https://doi.org/10.1523/JNEUROSCI.0210-15.2015, PubMed: 26224851

Chen, G., Pine, D. S., Brotman, M. A., Smith, A. R., Cox, R. W., Taylor, P. A., et al. (2022). Hyperbolic trade-off: The importance of balancing trial and subject sample sizes in neuroimaging. *Neuroimage*, *247*, 118786. https://doi.org/10.1016/j.neuroimage.2021.118786, PubMed: 34906711

Derrfuss, J., Brass, M., Neumann, J., & von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: Meta-analyses of switching and Stroop studies. *Human Brain Mapping*, *25*, 22–34. https://doi.org/10.1002/hbm.20127, PubMed: 15846824

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., et al. (2006). A core system for the implementation of task sets. *Neuron*, *50*, 799–812. https://doi.org/10.1016/j.neuron.2006.04.031, PubMed: 16731517

Dworetsky, A., Seitzman, B. A., Adeyemo, B., Neta, M., Coalson, R. S., Petersen, S. E., et al. (2021). Probabilistic mapping of human functional brain networks identifies regions of high group consensus. *Neuroimage*, *237*, 118164. https://doi.org/10.1016/j.neuroimage.2021.118164, PubMed: 34000397

Elliott, M. L., Knodt, A. R., & Hariri, A. R. (2021). Striving toward translation: Strategies for reliable fMRI measurement. *Trends in Cognitive Sciences*, *25*, 776–787. https://doi.org/10.1016/j.tics.2021.05.008, PubMed: 34134933

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146, PubMed: 17695343

Freund, M. C., Bugg, J. M., & Braver, T. S. (2021). A representational similarity analysis of cognitive control

during Color–Word Stroop. *Journal of Neuroscience*, *41*, 7388–7402. https://doi.org/10.1523/JNEUROSCI.2956-20.2021, PubMed: 34162756

Freund, M. C., Etzel, J. A., & Braver, T. S. (2021). Neural coding of cognitive control: The representational similarity analysis approach. *Trends in Cognitive Sciences*, *25*, 622–638. https://doi.org/10.1016/j.tics.2021.03.011, PubMed: 33895065

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., et al. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*, 326–340. https://doi.org/10.1037/dev0000075, PubMed: 26619323

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*, 201–225. https://doi.org/10.1037/0096-3445.137.2.201, PubMed: 18473654

Gaillard, A., Fehring, D. J., & Rossell, S. L. (2021). A systematic review and meta-analysis of behavioural sex differences in executive control. *European Journal of Neuroscience*, *53*, 519–542. https://doi.org/10.1111/ejn.14946, PubMed: 32844505

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., et al. (2017). Precision functional mapping of individual human brains. *Neuron*, *95*, 791–807. https://doi.org/10.1016/j.neuron.2017.07.011, PubMed: 28757305

Grady, C. L., Rieck, J. R., Nichol, D., Rodrigue, K. M., & Kennedy, K. M. (2021). Influence of sample size and analytic approach on stability and interpretation of brain–behavior correlations in task-related fMRI data. *Human Brain Mapping*, *42*, 204–219. https://doi.org/10.1002/hbm.25217, PubMed: 32996635

Gratton, C., Nelson, S. M., & Gordon, E. M. (2022). Brain–behavior correlations: Two paths toward reliability. *Neuron*, *110*, 1446–1449. https://doi.org/10.1016/j.neuron.2022.04.018, PubMed: 35512638

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*, 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060, PubMed: 19573611

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage*, *57*, 303–311. https://doi.org/10.1016/j.neuroimage.2010.12.027, PubMed: 21168515

Gustavson, D. E., Miyake, A., Hewitt, J. K., & Friedman, N. P. (2015). Understanding the cognitive and genetic underpinnings of procrastination: Evidence for shared genetic influences with goal management and executive function abilities. *Journal of Experimental Psychology: General*, *144*, 1063–1079. https://doi.org/10.1037/xge0000110, PubMed: 26389573

Haller, S. P., Kircanski, K., Stoddard, J., White, L. K., Chen, G., Sharif-Askary, B., et al. (2018). Reliability of neural activation and connectivity during implicit face emotion processing in youth. *Developmental Cognitive Neuroscience*, *31*, 67–73. https://doi.org/10.1016/j.dcn.2018.03.010, PubMed: 29753993

Hankin, B. L., Young, J. F., Abela, J. R., 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. *Journal of Abnormal Psychology*, *124*, 803–816. https://doi.org/10.1037/abn0000089, PubMed: 26595469

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166–1186. https://doi.org/10.3758/s13428-017-0935-1, PubMed: 28726177

Herzmann, G., & Curran, T. (2011). Experts' memory: An ERP study of perceptual expertise effects on encoding and recognition. *Memory & Cognition*, *39*, 412–432. https://doi.org/10.3758/s13421-010-0036-1, PubMed: 21264603

Infantolino, Z. P., Luking, K. R., Sauder, C. L., Curtin, J. J., & Hajcak, G. (2018). Robust is not necessarily reliable: From within-subjects fMRI contrasts to between-subjects comparisons. *Neuroimage*, *173*, 146–152. https://doi.org/10.1016/j.neuroimage.2018.02.024, PubMed: 29458188

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*, 825–841. https://doi.org/10.1006/nimg.2002.1132, PubMed: 12377157

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*, 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6, PubMed: 11516708

Kennedy, J. T., Harms, M. P., Korucuoglu, O., Astafiev, S. V., Barch, D. M., Thompson, W. K., et al. (2022). Reliability and stability challenges in ABCD task fMRI data. *Neuroimage*, *252*, 119046. https://doi.org/10.1016/j.neuroimage.2022.119046, PubMed: 35245674

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*, 154–169. https://doi.org/10.1016/j.tics.2017.11.006, PubMed: 29277256

Killgore, W. D. (2010). Effects of sleep deprivation on cognition. *Progress in Brain Research*, *185*, 105–129. https://doi.org/10.1016/B978-0-444-53702-7.00007-5, PubMed: 21075236

King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B., & Diedrichsen, J. (2019). Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nature Neuroscience*, *22*, 1371–1378. https://doi.org/10.1038/s41593-019-0436-x, PubMed: 31285616

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*, 229–235. https://doi.org/10.1016/j.tics.2007.04.005, PubMed: 17475536

Long, J. A. (2022). *Interactions: Comprehensive, user-friendly toolkit for probing interactions*. R package version 1.1.5. https://cran.r-project.org/package=interactions

MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of Experimental Psychology: General*, *121*, 12–14. https://doi.org/10.1037/0096-3445.121.1.12

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., et al. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*, 654–660. https://doi.org/10.1038/s41586-022-04492-9, PubMed: 35296861

Milham, M. P., & Banich, M. T. (2005). Anterior cingulate cortex: An fMRI analysis of conflict specificity and functional differentiation. *Human Brain Mapping*, *25*, 328–335. https://doi.org/10.1002/hbm.20110, PubMed: 15834861

Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., et al. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, *12*, 467–473. https://doi.org/10.1016/S0926-6410(01)00076-3, PubMed: 11689307

Muhle-Karbe, P. S., Andres, M., & Brass, M. (2014). Transcranial magnetic stimulation dissociates prefrontal and parietal contributions to task preparation. *Journal of Neuroscience*,

34, 12481–12489. https://doi.org/10.1523/JNEUROSCI.4931-13.2014, PubMed: 25209286

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., et al. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*, 115–126. https://doi.org/10.1038/nrn.2016.167, PubMed: 28053326

Pruim, R. H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage*, *112*, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064, PubMed: 25770991

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163. https://doi.org/10.2307/271063

Rypma, B., Berger, J. S., Prabhakaran, V., Bly, B. M., Kimberg, D. Y., Biswal, B. B., et al. (2006). Neural correlates of cognitive efficiency. *Neuroimage*, *33*, 969–979. https://doi.org/10.1016/j.neuroimage.2006.05.065, PubMed: 17010646

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., et al. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, *28*, 3095–3114. https://doi.org/10.1093/cercor/bhx179, PubMed: 28981612

Sheu, L. K., Jennings, J. R., & Gianaros, P. J. (2012). Test–retest reliability of an fMRI paradigm for studies of cardiovascular reactivity. *Psychophysiology*, *49*, 873–884. https://doi.org/10.1111/j.1469-8986.2012.01382.x, PubMed: 22594784

Silton, R. L., Heller, W., Towers, D. N., Engels, A. S., Spielberg, J. M., Edgar, J. C., et al. (2010). The time course of activity in dorsolateral prefrontal cortex and anterior cingulate cortex during top–down attentional control. *Neuroimage*, *50*, 1292–1302. https://doi.org/10.1016/j.neuroimage.2009.12.061, PubMed: 20035885

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*, 143–155. https://doi.org/10.1002/hbm.10062, PubMed: 12391568

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23(Suppl. 1)*, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051, PubMed: 15501092

Snyder, H. R., & Hankin, B. L. (2016). Spiraling out of control: Stress generation and subsequent rumination mediate the link between poorer cognitive control and internalizing psychopathology. *Clinical Psychological Science*, *4*, 1047–1064. https://doi.org/10.1177/2167702616633157, PubMed: 27840778

Snyder, H. R., Friedman, N. P., & Hankin, B. L. (2019). Transdiagnostic mechanisms of psychopathology in youth: Executive functions, dependent stress, and rumination. *Cognitive Therapy and Research*, *43*, 834–851. https://doi.org/10.1007/s10608-019-10016-z, PubMed: 31551642

Sober, E. (2002). Instrumentalism, parsimony, and the Akaike framework. *Philosophy of Science*, *69(Suppl. 3)*, S112–S123. https://doi.org/10.1086/341839

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. https://doi.org/10.1037/h0054651

Tang, R., Etzel, J. A., Kizhner, A., & Braver, T. S. (2021). Frontoparietal pattern similarity analyses of cognitive control in monozygotic twins. *Neuroimage*, *241*, 118415. https://doi.org/10.1016/j.neuroimage.2021.118415, PubMed: 34298081

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*, 242–249. https://doi.org/10.1016/j.psychres.2008.05.006, PubMed: 19564050

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*, 79–139. https://doi.org/10.1037/bul0000176, PubMed: 30596433

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*, 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x, PubMed: 26158964

Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, *8*, 254–274. https://doi.org/10.1037/1082-989X.8.3.254, PubMed: 14596490

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al. (2009). *Perspectives on Psychological Science*, *4*, 294–298. https://doi.org/10.1111/j.1745-6924.2009.01127.x, PubMed: 26158966

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*, 665–670. https://doi.org/10.1038/nmeth.1635, PubMed: 21706013