The organization and dynamics of corticostriatal pathways link the medial orbitofrontal cortex to future behavioral responses

Timothy D. Verstynen

J Neurophysiol 112:2457-2469, 2014. First published 20 August 2014; doi:10.1152/jn.00221.2014

J Neurophysiol 112: 2457–2469, 2014.
First published August 20, 2014; doi:10.1152/jn.00221.2014.

The organization and dynamics of corticostriatal pathways link the medial orbitofrontal cortex to future behavioral responses

Timothy D. Verstynen
Department of Psychology, Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania

Submitted 21 March 2014; accepted in final form 15 August 2014

Verstynen TD. The organization and dynamics of corticostriatal pathways link the medial orbitofrontal cortex to future behavioral responses. *J Neurophysiol* 112: 2457–2469, 2014. First published August 20, 2014; doi:10.1152/jn.00221.2014.—Accurately making a decision in the face of incongruent options increases the efficiency of making similar congruency decisions in the future. Contextual factors like reward can modulate this adaptive process, suggesting that networks associated with monitoring previous success and failure outcomes might contribute to this form of behavioral updating. To evaluate this possibility, a group of healthy adults ($n = 30$) were tested with functional MRI (fMRI) while they performed a color-word Stroop task. In a conflict-related region of the medial orbitofrontal cortex (mOFC), stronger BOLD responses predicted faster response times (RTs) on the next trial. More importantly, the degree of behavioral adaptation of RTs was correlated with the magnitude of mOFC-RT associations on the previous trial, but only after accounting for network-level interactions with prefrontal and striatal regions. This suggests that congruency sequencing effects may rely on interactions between distributed corticostriatal circuits. This possibility was evaluated by measuring the convergence of white matter projections from frontal areas into the striatum with diffusion-weighted imaging. In these pathways, greater convergence of corticostriatal projections correlated with stronger functional mOFC-RT associations that, in turn, provided an indirect pathway linking anatomical structure to behavior. Thus distributed corticostriatal processing may mediate the orbitofrontal cortex's influence on behavioral updating, even in the absence of explicit rewards.

adaptation; congruency; corticostriatal processing; diffusion-weighted imaging; fMRI

A CRITICAL FEATURE of adaptive decision-making is the ability to modify future actions based on the success or failure of previous decisions. Consider for a moment a car spinning on a patch of ice. To stop the spin, the driver must suppress the automatic urge to turn the steering wheel against the direction of spin and instead turn the wheel into the spin. Successfully stopping the out-of-control car increases the likelihood of the driver making the right decision the next time he hits a patch of ice further down the road. This rapid learning reflects a form of updating that is classically known as conflict adaptation (Botvinick et al. 2001; Gratton et al. 1992), where successful resolution of conflicting response cues (e.g., automatic desire to turn against the spin vs. correct response of turning into the spin) increases the efficiency of resolving similar conflicts in the future.

Rapid updating of responses after a cue conflict reflects the amalgamation of many different cognitive processes (Egner 2007). For example, in congruency paradigms like the Stroop, Simon, and Flanker tasks, subjects can learn to associate distractor stimuli with congruent target stimuli more than with any incongruent stimuli because of the regularity of their co-occurrence (Schmidt and De Houwer 2011). In this way repeated congruent trials (i.e., trials without conflicts between target and distractor) are executed faster through contingency learning. Feature integration may also contribute to updating after repeated trial types since the repetition of stimulus features may prime the perceptual identification of target stimuli and thus speed subsequent response selection (Hommel 2004; Mayr et al. 2003). Finally, adaptation to repeated cue conflicts can occur through adaptation to conflict monitoring processes by biasing attention toward or away from target-relevant stimulus features (Botvinick et al. 2001; Kim and Cho 2014; Schmidt and Weissman 2014). In this case, resolving a conflict in response cues facilitates attention toward the relevant stimulus features and suppresses attention to the nonrelevant (i.e., distracting) stimulus features on the following trial. Given the many possible mechanisms that could explain trial-by-trial plasticity in most types of cue-conflict paradigms, this form of adaptation is generally referred to as the congruency sequencing effect (CSE) or sometimes the Gratton effect.

The plurality of processes linked to the CSE suggests that it relies on a broad and distributed network of brain regions. Studies premised on conflict monitoring theory (Botvinick et al. 2001) usually highlight the role of the dorsal anterior cingulate cortex (dACC) and dorsolateral prefrontal cortex (DLPFC) in conflict processing. Neuroimaging and electrophysiological experiments have shown that signal changes in both the dACC and DLPFC are modulated on conflict trials when the previous trial also had a response conflict (Casey et al. 2000; Kim et al. 2013, 2014; Sheth et al. 2011, 2012). However, the dACC and DLPFC are not the only regions that respond to incongruent response cues. Activity in the superior frontal gyrus (SFG), superior parietal lobe, and cerebellum have all been associated with stimulus incongruency, and this activity is thought to be linked to violations of expectations in visuospatial attention (Casey et al. 2000). Basal ganglia areas like the caudate nucleus (Casey et al. 2000; Watanabe and Munoz 2009) and the subthalamic nucleus (Brittain et al. 2012) have also been associated with congruency processing on the current trial and are believed to reflect the competition of potential action plans. While it is clear that these other frontal, parietal, and basal ganglia regions are engaged when resolving an ongoing response conflict, it remains unclear to what extent these other regions might contribute to adaptation after a stimulus incongruency.

Address for reprint requests and other correspondence: T. D. Verstynen, Carnegie Mellon Univ., Dept. of Psychology and Center for the Neural Basis of Cognition, 342C Baker Hall, Pittsburgh, PA 15213 (e-mail: timothyv@andrew.cmu.edu).

Knowing the brain networks that contribute to the CSE can provide insights into the mechanisms that mediate this form of rapid plasticity. For example, while traditional hypotheses of the CSE emphasize a form of temporal association across trials (Botvinick et al. 2001; Schmidt and De Houwer 2011; Ullsperger et al. 2005), the observation of activity in basal ganglia regions during congruency processing (Brittain et al. 2012; Casey et al. 2000; Watanabe and Munoz 2009) hints at a possible role of reinforcement learning in the CSE. Indeed, consistent with this idea, recent behavioral studies have found that the CSE is modulated by contextual information, including reward signals (Braem et al. 2012; van Steenbergen et al. 2009). For example, introducing financial rewards for successful resolution of conflicting cues can lead to greater behavioral changes on the next trial, and this effect is enhanced in subjects with greater reward sensitivity (Braem et al. 2012), indicating that part of the CSE may be modulated by feedback signals linked to monitoring the reinforced outcomes of previous trials.

Mechanistically, lesion studies (Chudasama and Robbins 2003) and neural network simulations (Frank and Claus 2006) suggest that outcome monitoring is regulated, in part, by rostral corticostriatal circuits. Specifically, these models of behavioral updating posit that action selection signals from lateral frontal areas are modulated by the recent history of successes and failures that are being monitored by the orbitofrontal cortex (Frank and Claus 2006). Frontostriatal pathways leading to successful decisions are thought to be upregulated, while pathways leading to unsuccessful or inappropriate decisions are downregulated, via feedback processes. If this type of outcome assessment is happening during the CSE, then it should be possible to detect signatures of the updating process in the dynamics and structure of the frontal corticostriatal circuits.

Anatomically, there is a growing body of evidence for convergent projections from orbitofrontal (e.g., outcome monitoring), lateral prefrontal (e.g., response selection), and medial prefrontal (e.g., conflict monitoring) cortical areas into the striatum. Tracing studies in nonhuman primates have reported diffuse projections from orbitofrontal regions that terminate in dorsal striatal regions that contain dense projections from lateral and medial prefrontal areas (Averbeck et al. 2014; Haber et al. 1995, 2006). Indeed, the overlap of frontal afferents has been postulated as one mechanism for integrating reward and executive information in basal ganglia pathways (Haber and Knutson 2010). This integration hypothesis of corticostriatal loops postulates that within a particular basal ganglia loop information is independent and segregated from the other parallel loops (Alexander et al. 1986); however, diffuse overlap of inputs from different cortical sources is one of several mechanisms that allow for sharing information across these otherwise segregated systems. If integration across corticostriatal circuits is critical for the CSE, then greater overlap of frontostriatal anatomical projections should predict more efficient adaptation following a response conflict.

The present study set out to evaluate whether the functional dynamics of frontal corticostriatal circuits correlates with behavioral response adaptation following a cue conflict and whether this brain-behavior relationship could be explained by the anatomical organization of the underlying white matter pathways. To evaluate this, a sample of neurologically healthy adults performed the color-word version of the Stroop task (Stroop 1935) while event-related brain dynamics were measured with functional MRI (fMRI). Subsequent imaging using a high-angular resolution form of diffusion-weighted imaging allowed for topographic mapping of corticostriatal white matter projections in the same subjects. Combining both functional and structural imaging approaches allows for identification of the network-level properties underlying behavioral updating and possible insights into the learning mechanisms supporting the CSE.

MATERIALS AND METHODS

Participants. Twenty male and ten female subjects were recruited from the Pittsburgh area and the Army Research Laboratory in Aberdeen, Maryland. All subjects were neurologically healthy, with no history of either head trauma or neurological or psychiatric illness. Subject ages ranged from 21 to 45 yr at the time of scanning (mean age of 31 yr), and four subjects were left-handed (2 men, 2 women). All participants gave written informed consent to participate in protocols reviewed and approved by Carnegie Mellon University Institutional Review Board and conforming with the Declaration of Helsinki and were financially remunerated for their participation.

Stroop task. Participants performed the color-word version of the Stroop task (Botvinick et al. 2001; Gratton et al. 1992; Macleod 1991; Stroop 1935) comprised of congruent, incongruent, and neutral conditions while in the MR scanner. Participants were instructed to ignore the meaning of the printed word and respond to the ink color in which the word was printed. For example, in the congruent condition the words "RED," "GREEN," and "BLUE" were displayed in the ink colors red, green, and blue, respectively. In this condition, attentional demands were low because the ink color matched the prepotent response of reading the word, so response conflict was at a minimum. However, for the incongruent condition the printed words were different from the ink color in which they were printed (e.g., the word "RED" printed in blue ink). This condition elicited conflict because responding according to the printed word would result in an incorrect response. As a result, attentional demands were high and participants needed to inhibit the prepotent response of reading the word and respond according to the ink color in which the word was printed. On the other hand, the neutral condition consisted of noncolor words presented in an ink color (e.g., the word "CHAIR" printed in red ink) and had a low level of conflict and low attentional demands.

Participants were instructed to respond to the ink color in which the text appeared by pressing buttons under the index, middle, and ring fingers on their right hand, each button corresponding to one of the three colors (red, green, and blue, respectively) on an MR-safe response box. The task was briefly practiced in the scanner to acquaint the participant with the task and to ensure understanding of the instructions. The task began with the presentation of a fixation cross hair for 1,000 ms followed by the Stroop stimulus for 2,000 ms, during which participants were instructed to respond as quickly as possible. The interstimulus interval between successive trial starts was sampled from an exponential distribution, between 3 and 20 s with a mean of 4 s and a median of 3 s, in order to ensure accurate deconvolution of the hemodynamic response. Condition types were pseudorandomized in an event-related fashion. A total of 120 trials were presented to each participant (42 congruent, 42 neutral, 36 incongruent). A lower number of incongruent trials was used in order to reduce the expectancy of a stimulus conflict relative to the other conditions. Stimuli were back-projected onto a screen located at the back of the MR bore with an MR-safe projector. Participants viewed stimuli by using a mirror attached to the top of the head coil. If necessary, vision was corrected to at least 20/40 with MR-safe plastic glasses and corrective lenses.

Behavioral analysis. The primary behavioral variable of interest was response time (RT), recorded as the time between cue onset and

registered key press (in milliseconds). All first-level analyses were restricted to correct responses. To determine condition-level effects, a one-way repeated-measures ANOVA was used, as were post hoc one-sample *t*-tests.

To measure the CSE, the vector of RTs was first mean-centered to remove baseline biases in overall response speed across subjects. Next, all incongruent trials were selected, starting with the second trial of the series. Then these trials were categorized by the condition label of the preceding trial. Neutral trials were excluded from this comparison. The CSE was then calculated for each subject by subtracting the mean RT of incongruent trials preceded by a congruent condition (μ_{CI}) from that of trials preceded by an incongruent condition (μ_{II}), i.e., adaptation = $\mu_{II} - \mu_{CI}$. Higher values reflect greater adaptation after repeated incongruent trials. Significance of the CSE was determined with a one-sample *t*-test on these adaptation scores. This analysis was performed with custom-written MATLAB scripts (Release 2012b, The MathWorks, Natick, MA).

MRI acquisition. All 30 participants were scanned on a Siemen's Verio 3T system in the Scientific Imaging & Brain Research (SIBR) Center at Carnegie Mellon University with a 32-channel head coil. A high-resolution (1-mm^3 voxel) T1-weighted brain image was acquired for all participants consisting of 176 contiguous slices with a magnetization-prepared rapid gradient echo imaging (MPRAGE) sequence. A blood oxygenation level-dependent (BOLD) contrast with echo planar imaging (EPI) sequence was used for all functional MRI (TE = 20 ms, TR = 1,500 ms, flip angle = 90°). Thirty contiguous slices (3.2 mm × 3.2 mm × 4 mm) were collected in an ascending and sequential fashion parallel to the anterior and posterior commissures. A 50-min, 257-direction diffusion spectrum imaging (DSI) scan was collected after the fMRI sequences, with a twice-refocused spin-echo EPI sequence and multiple b values (TR = 9,916 ms, TE = 157 ms, voxel size = 2.4 mm^3, FoV = 231 × 231 mm, b-max = 5,000 s/mm^2, 51 slices). Minimization of head motion during acquisition was done through a custom-designed setup of foam padding within the coil, designed to minimize variance of head motion along the pitch and yaw rotation directions. This setup also included a chin restraint that held the participant's head to the receiving coil itself. Preliminary work on EPI images at the imaging center showed that this setup minimized resting head motion to ~1-mm maximum deviation for most subjects.

fMRI analysis. Two participants were excluded from all fMRI analyses, one because of an error in writing the data files to disk and another because of an error in the acquisition process (incorrect scan sequence used). This left a total sample size of 28 for functional imaging analysis. Functional data from each participant were processed and analyzed with the SPM8 toolbox. Prior to analysis, the EPI images for each participant were realigned to the first image in the series and corrected for differences in the slice acquisition time. All images were then coregistered to MNI-space with a nonlinear spatial normalization approach (ICBM-152 space template regularization, 16 nonlinear iterations). These images were then smoothed with a 4-mm isotropic Gaussian kernel.

Estimates of task-related responses at each voxel were determined with a reweighted least-squares generalized linear model (GLM) approach (Diedrichsen and Shadmehr 2005) that minimizes the influence of movement-related noise in the signal. Only responses on correct trials were included in the analysis. Each trial onset was convolved with a double-gamma hemodynamic response function, with each Stroop condition (Congruent, Incongruent, and Neutral) entered as a separate explanatory variable. For identification of areas with responses associated with each trial type, a condition-wise, whole-brain, random-effects analysis was performed. Each Stroop condition type was estimated as a separate independent variable, providing a map of regression coefficients (β) for each condition at every voxel. To isolate incongruent trial-related areas, a contrast difference between incongruent and neutral trials ($\beta_{Inc} - \beta_{Net}$) was calculated and a one-sample *t*-test across all subjects was used to determine significance at each voxel. The statistical threshold for significant effects was determined with a false discovery rate (Chumbley and Friston 2009) across all gray matter voxels of 0.05 ($q < 0.05$). Clusters of >20 contiguously active voxels were then kept for subsequent region of interest (ROI) analyses.

After the condition-specific analysis, a single-trial event-related analysis was performed (Rissman et al. 2004). This followed the same analytical procedures as described above, with the exception that each individual trial was included as a separate independent variable in the GLM, providing a regression coefficient for each trial. The average single-trial response of each ROI identified in the previous analysis was estimated by averaging the single-trial regression coefficients across all voxels in an ROI.

Indirect pathway analysis. Statistical mediation was performed with a nested regression and permutation-based statistical inference approach (Preacher and Hayes 2008) with the Bootstrap Regression Analysis of Voxelwise Observations (BRAVO) toolbox (**https://sites.google.com/site/bravotoolbox**). These models estimate the indirect pathways that link trial-by-trial variation in BOLD response to variation in single-trial RTs. Based on the effect sizes of single-trial BOLD estimates from previous studies (Rissman et al. 2004), as well as the estimated effect size of BOLD-RT relationships (Weissman and Carp 2013), a minimum of 100 trials would be needed for reliable model estimation per subject (Mackinnon et al. 2002). Therefore this analysis was collapsed across all three trial conditions. In addition, ROIs were selected for the indirect pathway analysis only if their single-trial BOLD responses met two criteria: *1*) they were correlated with the current-trial RT and *2*) they were correlated with single-trial BOLD responses in the cluster found on the gyrus rectus (see RESULTS).

In the model, the vector of RTs was the dependent variable (Y), the vector of single-trial regression coefficients from the medial orbitofrontal cortex (mOFC) was the independent variable (X), and the single-trial regression coefficients from each of the associated ROIs were included as candidate mediating pathways (M). The selection of the independent, mediating, and dependent variables was based on empirical observations (see RESULTS). All variables were mean centered prior to analysis, and pathway coefficients shown in *Eqs. 1–3* were estimated with an ordinary least-squares regression.

$$Y = cX + \eta \quad (1)$$

$$M = aX + \eta \quad (2)$$

$$Y = c'X + bM + \eta \quad (3)$$

The total (c) pathway estimates the simple relationship between X and Y, without the inclusion of mediating variables (*Eq. 1*). The indirect (a*b) pathway via each mediating variable is estimated by computing how much the X variable predicts the candidate mediator (a pathway, *Eq. 2*) and the influence the mediator variable has on the Y variable (b pathway) when taking into account the relationship between the X and Y variables (*Eq. 3*). Finally, the direct pathway (c') reflects the residual relationship between X and Y after accounting for the influence of the mediating, indirect pathway. The η term in each equation reflects the residual noise in the estimator. This is assumed to be Gaussian and temporally independent (i.e., white noise) across trials.

A permutation approach was used to estimate the significance of the pathways in each model. For each iteration of the algorithm, the values in the variable vectors (X, Y, and M) were scrambled independently and *Eqs. 1–3* reestimated. The values for a, b, c', and a*b from these permuted models were stored in a separate matrix, and this process was repeated for 1,000 iterations per model. The significance of the direct and indirect paths was determined from the distribution of permuted values with a bias-corrected and accelerated method (Diciccio and Efron 1996). Statistical significance was determined after adjusting for multiple comparisons with a false discovery rate (q) of 0.05.

A second control model was also run that included dummy regressors as a control for trial type. For this analysis, an $N \times 3$ binary

matrix was included as a covariate in the model, where N is the number of trials and each column represents one of the three trial types. All other aspects of the model were the same as those described in the previous two paragraphs.

Diffusion MRI reconstruction. All DSI images were processed with q-space diffeomorphic reconstruction (Yeh and Tseng 2011) implemented in DSI Studio (**http://dsi-studio.labsolver.org**). The normalization to template space was conducted with a nonlinear spatial normalization approach (Ashburner and Friston 1999), and a total of 16 iterations were used to obtain the spatial mapping function between the individual subject diffusion space, i.e., the map of quantitative anisotropy (QA) values, and the FMRIB 1-mm template fractional anisotropy atlas in MNI space. QA is a diffusion anisotropy metric similar to the more common fractional anisotropy (FA) index in diffusion-weighted imaging (Yeh et al. 2010). From here orientation distribution functions (ODFs) were reconstructed to spatial resolution of 2 mm^3 and a diffusion sampling length ratio of 1.25. To determine the average tractography space, a template image was generated that was composed of the average whole-brain ODF maps across all 30 subjects.

Fiber tractography. All tractography was performed with DSI Studio (November 14, 2012 build). Streamlines were generated with a generalized deterministic tractography algorithm (Yeh et al. 2013). Tractography (see Fig. 5) was performed between pairs of ROI masks selected from the SRI24 multichannel atlas (Rohlfing et al. 2010) and a mask of the striatum generated by merging the caudate nucleus and putamen masks and expanding this mask by 1 voxel (2 mm). Cortical targets were selected in order to encompass most frontal areas, as well as portions of the basal ganglia network. This was done by selecting 11 frontal ROI masks from the SRI24 atlas: gyrus rectus (Rectus), ventral medial prefrontal cortex (Frontal_Med_Orb), the lateral and middle orbitofrontal gyri (Frontal_Mid_Orb and Frontal_Inf_Orb), segments of the inferior frontal gyrus (operculum, Frontal_Inf_Oper; triangularus, Frontal_Inf_Tri), insula (Insula), middle frontal gyrus (Frontal_Mid), lateral SFG (Frontal_Sup), medial SFG (Frontal_Sup_Medial), and anterior cingulate (Cingulum_Ant).

For each cortical ROI, the tracking session started with seed positions that were randomly started anywhere within the brain mask and fiber progression that started in opposite directions from a random initial orientation. Fiber progression continued with a step size of 1 mm, and at each step the next directional estimate of each voxel was weighted by 80% of the previous moving direction and 20% by the incoming direction of the fiber. This continued until the underlying QA index dropped below 0.15 or necessitated a turn of $>75°$. This process was repeated for 31,100,100 seeds (approximating 300 samples per voxel in the brain). Streamlines were included in the final data set only if they met the following criteria: *1*) the streamline had a length of <90 mm [10 mm above the maximum length of frontostriatal fibers based on previous work (Verstynen et al. 2012)] and *2*) one end of the streamline terminated in the striatum mask and the other end of the streamline terminated in the cortical ROI mask.

Structural overlap analysis. The main focus of the tractography analysis was to estimate convergence of cortical projections into the striatum (see RESULTS for motivation). For this, the topology of projections from each cortical system to the striatum mask was determined. For every subject and cortical ROI (see *Fiber tractography*), the percentage of streamlines terminating in the cluster of striatal voxels was calculated. Significance at each cortical ROI was determined with a one-sample t-test calculated across subjects, with an uncorrected threshold of $P < 0.005$. For simplicity of visual presentation (see Fig. 5), the cortical ROIs were organized into four cluster sets: orbitofrontal (Frontal_Mid_Orb, Frontal_Inf_Orb, Rectus), medial frontal (Frontal_Med_Orb, Cingulum_Ant, Frontal_Sup_Medial), lateral frontal (Frontal_Mid, Frontal_Sup, Frontal_Inf_Oper, Frontal_Inf_Tri), and insular cortex (Insula).

To measure the amount of overlapping projections into each voxel, an overlap index (OI) was calculated for each subject (s). This determines the percent overlap of streamlines, from different ROIs, into the same striatum voxel.

$$\mathrm{OI}_s = \frac{1}{N_{\mathrm{ROIs}s}^2 - N_{\mathrm{ROIs}s}} \sum_{i=1}^{N_{\mathrm{ROIs}s}-1} \sum_{j-i+1}^{N_{\mathrm{ROIs}s}} \left(\frac{1}{N_{\mathrm{vox}}} \sum_{v=1}^{N_{\mathrm{vox}}} P_v(f_i \cap f_i) \right) \quad (4)$$

For all N_{vox} voxels in the striatum mask, the conditional probability, $P_v(f_i \cap f_j)$, that at least one streamline from any pair of cortical ROIs, f_i and f_j, terminates within the voxel was determined. If an overlap of streamlines was detected, then $P_v(f_i \cap f_j) = 1$; otherwise $P_v(f_i \cap f_j) = 0$. This was averaged across all striatal voxels to provide the probability of overlap of projections for any given pair of cortical ROIs. This process was repeated for all pairs of cortical ROIs and averaged to create a composite index for each subject. The OI was calculated independently for the left and right hemisphere networks. All analyses of fiber streamlines were performed with custom-written MATLAB routines.

RESULTS

Behavioral congruency effects. Behavioral responses in the scanner were consistent with previous studies (Botvinick et al. 2001; Gratton et al. 1992; Macleod 1991; Stroop 1935). On all correct trials, there was a main effect of stimulus condition on RTs [Fig. 1*A*; $F(58,2) = 67.88$, $P < 0.001$]. Compared with neutral control trials, subjects were slower to respond when there was a cue conflict [incongruent trials; $t(27) = 10.02$, $P < 0.0001$] and faster on trials with redundant cues [congruent trials; $t(27) = -5.68$, $P < 0.0001$]. Responses during incongruent trials were modulated depending on the structure of the previous trial. When the previous trial was also incongruent participants were 32 ms faster to respond than when the previous trial was congruent [Fig. 1*B*; $t(27) = 2.82$, $P = 0.004$]. Participants were slightly more accurate (2% fewer errors) on incongruent trials when the previous trial was also incongruent [Fig. 1*C*; $t(27) = 2.05$, $P = 0.025$]. A similar repetition effect was observed on congruent trials, but attenuated. On average, participants were 19 ms faster on a congruent trial when the previous trial was also congruent than when the previous trial was incongruent [Fig. 1*B*; $t(27) = -11.22$, $P < 0.0001$]. No improvement to the error rate was observed for repeating congruent trials [$t(27) = 0.41$, $P = 0.34$]; however, this may be due to a low base rate of errors on congruent trials to begin with ($\sim$1%).

Networks associated with response conflict. For behavioral updating to take place, a critical computation has to occur on or immediately after the preceding trial. To first isolate regions associated with processing stimulus incongruency, a whole-brain random effects analysis was conducted that identified regions selectively responsive to incongruent trials compared with neutral trials. This contrast was performed to reduce the number of regions being analyzed, relative to a whole-brain voxelwise approach, by selecting only the network of regions that are selectively responsive, in some way, to aspects of incongruency processing. Such analysis is sensitive to conflict monitoring processes but is also sensitive to other functions related to incongruency processing, including time on task and error anticipation (Grinband et al. 2011; Weissman and Carp 2013). Twenty distributed regions were associated with task-specific responses during incongruent trials, when compared against neutral trials (Fig. 2, *A* and *B*; Table 1). Many of these

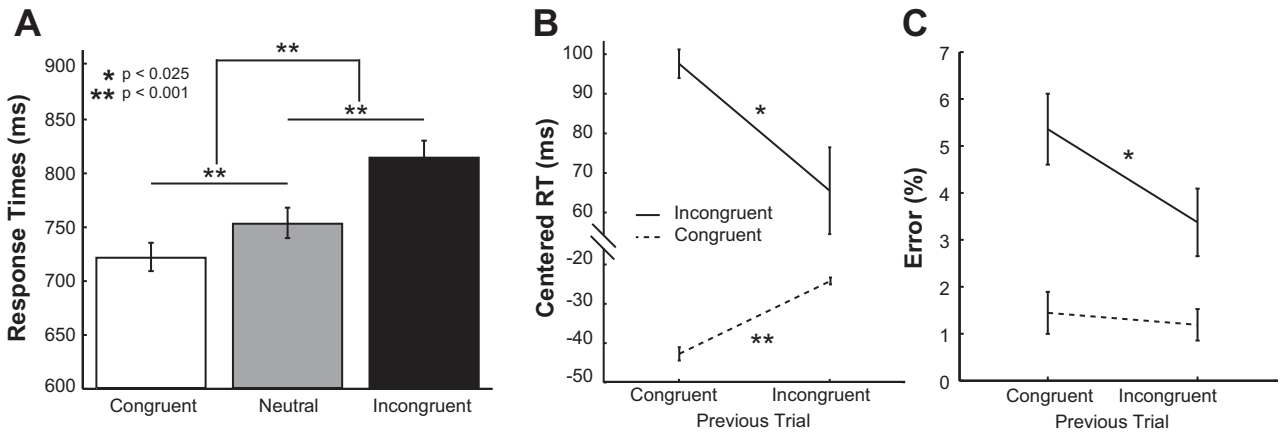

Fig. 1. Behavioral responses during the color-word Stroop task. *A*: in general, response times (RTs) for accurate incongruent trials were slower than for neutral trials, while RTs for congruent trials were faster than for neutral trials. *B*: incongruent trials preceded by incongruent trials had faster responses than when they were preceded by a congruent stimulus. In contrast, congruent trials that were preceded by an incongruent stimulus were slower than when 2 congruent trials were repeated. All RTs were mean centered prior to analysis. *C*: the advantage for condition repetition was also seen in accuracy, with repetition of incongruent trials leading to fewer errors on the second trial. This effect, however, was not seen for congruent trials. All error bars show SE across subjects.

have been previously associated with incongruency in the Stroop task (Banich et al. 2000), as well as other cue conflicts in other tasks (Casey et al. 2000; Weissman and Carp 2013).

One region of particular interest was a cluster of voxels on the gyrus rectus of the mOFC that had a stronger BOLD response during incongruent trials than neutral or congruent trial types (Fig. 2*A*). This mOFC cluster was one of two clusters identified in the whole-brain analysis that had a negative contrast value (see Table 1). More importantly, this aspect of the orbitofrontal cortex is thought to be associated with monitoring recent trial outcomes to modulate future decisions (Frank and Claus 2006). To better understand the nature of this negative contrast effect in the mOFC, the average trial-evoked BOLD response for each stimulus condition was

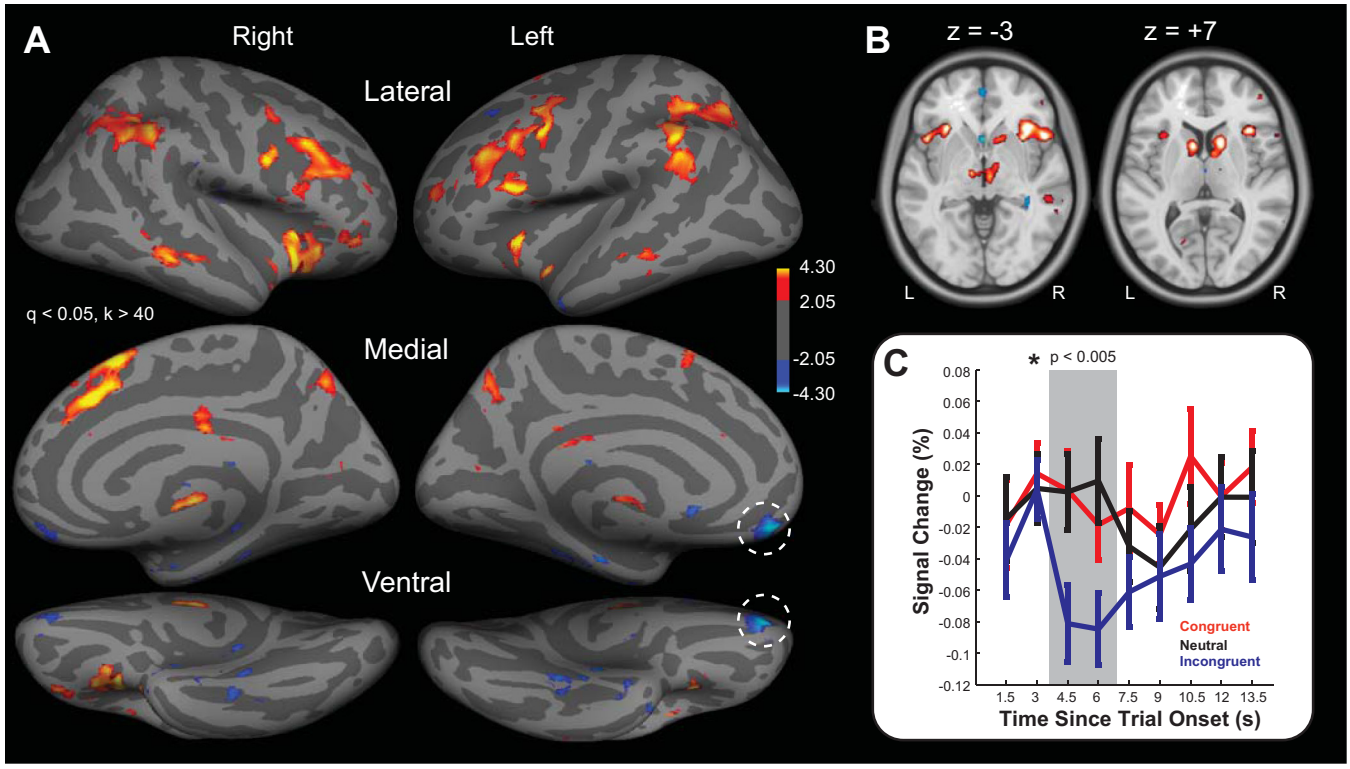

Fig. 2. Stroop-related blood oxygenation level-dependent (BOLD) responses. *A*: activation maps on an inflated brain showing voxels with differential BOLD responses to cue incongruency (i.e., incongruent vs. neutral trials). Warm colors indicate areas where the incongruent trial response was more positive than the neutral trial response. Cool colors indicate areas where the neutral trial response was more positive. Activation maps are adjusted for multiple comparisons with a false discovery rate [FDR (q)] of <0.05 and restricted to clusters of 20 or more continuously connected voxels. Dashed circles highlight the negative cluster on the gyrus rectus of the medial orbitofrontal cortex (mOFC). *B*: 2 axial slices showing the same activation patterns as in *A*. Slice position in MNI coordinates is shown above each slice. *C*: trial-locked BOLD responses from the mOFC cluster for each task condition. Averaged signal change values in the peak of the BOLD response (gray bar) show a significant modulation based on condition type [repeated measures $F(2,54) = 6.61$, $P = 0.003$]. Error bars show SE across subjects.

Table 1. *Clusters with incongruent trial-related responses compared with neutral trials*

ROI	N_{Vox}	MNI Coordinates (*x*,*y*,*z*)	Peak *t*	Peak *P*
L_Caudate	93	−10, 10, 8	5.55	<0.0001
L_IPS	712	−54, −44, 38	5.69	<0.0001
L_Insula	267	−32, 22, 0	7.11	<0.0001
L_MFG	842	−48, 18, 28	7.77	<0.0001
L_MTG	41	−56, −34, −8	4.86	<0.0001
L_PHC	56	−14, −12, −18	−3.62	0.0006
L_mOFC	87	−10, 52, −16	−3.61	0.0006
L_TPJ	64	−58, −50, 28	4.70	<0.0001
L_STN	26	−10, −14, −2	5.06	<0.0001
Precuneus	98	8, −68, 44	4.35	<0.0001
R_Medial SFG	819	6, 16, 60	6.70	<0.0001
R_Caudate	223	14, 12, 4	6.06	<0.0001
R_MFG (ventral)	45	48, 40, −8	4.94	<0.0001
R_IPS	716	58, −54, 40	5.04	<0.0001
R_Insula	742	30, 18, −12	8.26	<0.0001
R_MFG	736	50, 16, 28	6.21	<0.0001
R_MTG	68	64, −30, −10	4.86	<0.0001
R_PCS	97	48, 12, 54	5.69	<0.0001
R_STN	73	6, −16, −2	5.15	<0.0001
R_SMG	52	62, −48, 20	4.93	<0.0001

MNI coordinates, *t*-statistics, and *P* values are for the peak voxel in the cluster. IPS, intraparietal sulcus; MFG, middle frontal gyrus; mOFC, medial orbitofrontal cortex; TPJ, temporal-parietal junction; STN, subthalamic nucleus; SFG, superior frontal gyrus; PCS, precentral sulcus; SMG, supramarginal gyrus; MTG, middle temporal gyrus; PHC, parahippocampal cortex.

estimated (Fig. 2*C*). Unlike congruent and neutral trials, the mOFC exhibited a clear time-locked response to incongruent trials, reflected as a dip in the BOLD signal. Therefore, this region was responsive to conflict in response cues, but with a negatively directed evoked BOLD signal.

Associations with current and future trial responses. The whole-brain analysis identified several regions selective to stimulus incongruency. To understand how activity in these areas might relate to trial-by-trial variation in behavioral performance, a single-trial regression analysis was adopted to measure the evoked BOLD responses on individual trials (see MATERIALS AND METHODS). This trialwise BOLD response measure was then correlated with behavioral RTs, across all conditions, for each incongruency-related cluster (Fig. 3*A*). Of the 20 ROIs identified, trialwise BOLD responses in 15 regions were significantly correlated with RTs on the current trial. The direction of these correlations is consistent with previous reports that the magnitude of the BOLD response in many conflict-related areas may reflect the longer time on task during incongruent trials (Weissman and Carp 2013).

However, unlike the other ROIs, the mOFC cluster did not exhibit a simple correlation with current-trial RTs, suggesting that its responses are not directly predictive of current-trial response speeds. Instead, the mOFC cluster was strongly correlated with the RT on the following trial (Fig. 3*B*). The direction of this association implies that more negative (i.e., stronger) mOFC responses on the current trial predict slower responses on the next trial. After controlling for the influence of trial type on RT, with a partial correlation analysis, this association between mOFC responses and future responses remains significant but changes direction {mean $r = -0.046$, 95% confidence interval (CI) = $[-0.022, -0.071]$}. This directional change in the partial correlation reflects the fact that accounting for trial type on RT flips the overall direction of the

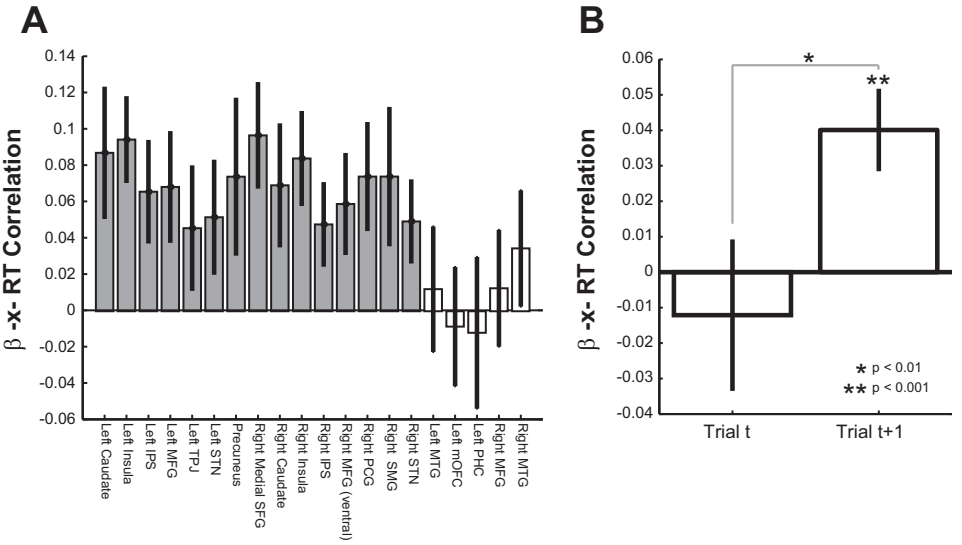

Fig. 3. Brain-behavior correlations. *A*: correlation coefficients for the association between RT on the current trial and the single-trial BOLD estimate from 20 regions of interest (ROIs). Filled bars indicate ROIs with significant trialwise associations across subjects (FDR corrected, $q < 0.05$). Error bars show the Bonferroni-corrected 95% confidence interval (CI) across subjects. *B*: correlation for the mOFC responses against current-trial (*trial t*) RT, showing the same data as in *A*, and the next-trial RT (*trial t + 1*). Plotting convention as in *A*.

RT vector and thus causes the direction of subsequent brain-behavior associations to flip. Nonetheless, the inferential direction of the association between mOFC responses and next-trial RT remains the same. A consistent, but weaker, lag-1 correlation with RT was also found for the right middle frontal gyrus (MFG) ($r = -0.027$) and medial wall SFG ($r = -0.024$). Within the mOFC cluster, this association with future responses is consistent with the hypothesis that the orbitofrontal cortex may play a role in updating future responses based on current trial outcomes (Frank and Claus 2006).

Distributed networks linking orbitofrontal responses to behavior. At first glance, finding that mOFC activity was not correlated with current-trial response speed may appear contradictory to the hypothesis that ventral corticostriatal systems are involved in the CSE process. After all, in order to contribute to learning an evaluation of current trial outcomes is necessary to modify future responses (Dayan and Abbott 2001). However, given the large and distributed network of areas associated with incongruency in the previous analysis, it is possible that a residual association between mOFC and current-trial RT is masked by indirect pathways linking orbitofrontal responses to current trial outcomes.

To test this hypothesis, a nested regression analysis was used to identify indirect pathways linking gyrus rectus activity to current trial responses (see MATERIALS AND METHODS; Preacher and Hayes 2008). For this analysis, the mOFC ROI was the independent variable, current-trial RT was the dependent variable, and all remaining ROIs that correlated with current-trial RT (see Fig. 3*A*) were tested as possible indirect pathways with a multiple mediation model. Three bilateral clusters were found to be indirect pathways linking mOFC responses to RT on the current trial (Fig. 4*A*): the MFG, the insula, and caudate nucleus. Thus there were dorsolateral, ventrolateral, and striatal regions that statistically mediated the association between mOFC responses on the current trial and the speed of behavioral responses on that same trial.

To show that this result is not biased by the ROI selection process, which itself was determined by differences in trial responses that are also correlated with differences in RT, the indirect pathway analysis was repeated for these six regions with a model that also controlled for trial type. All three bilateral ROIs were still found to be statistically significant indirect pathways after controlling for trial type (Left Caudate: mean $a*b = 1.55$, 95% CI = 0.49–2.62, $P = 0.004$; Right Caudate: mean $a*b = 1.84$, 95% CI = 0.36–3.37, $P = 0.011$; Left MFG: mean $a*b = 1.91$, 95% CI = 0.69–3.14, $P = 0.0025$; Right MFG: $a*b = 1.20$, 95% CI = 0.31–2.09, $P = 0.0066$; Left Insula: mean $a*b = 1.41$, 95% CI = 0.47–2.35, $P = 0.0034$; Right Insula: mean $a*b = 1.62$, 95% CI = 0.75–2.49, $P = 0.0005$). This analysis shows that the indirect pathways linking mOFC responses to current-trial RT cannot be explained by the main effect of trial condition.

By nature of the nested regression analysis (see *Eq. 3* in MATERIALS AND METHODS), the direct pathway (c') represents the residual relationship between mOFC responses and RT, after accounting for the indirect interactions with the mediating ROIs, i.e., $c'X = Y - bM - \eta$. Averaged across all significant indirect pathways, there was a consistent negative c' pathway between mOFC and behavior [Fig. 4*B*; $t(27) = -3.90$, $P = 0.0002$] (Fig. 3*B*). This effect was also significant for the residual from each indirect pathway when looked at individually (all $P < 0.0002$). Taking into account the direction of the evoked BOLD response in the gyrus rectus (Fig. 2*C*), this means that stronger (i.e., more negative) evoked responses in the mOFC are associated with slower reaction times on the current trial, after controlling for indirect associations with the lateral prefrontal cortex, caudate nucleus, and insula. Thus the mOFC-RT association on the current trial is normally obscured by the interactions between the indirect regions that also relate to both mOFC responses and behavior.

To understand what, if any, predictive value this direct pathway between mOFC responses and current-trial RT has on adaptation, an individual differences analysis was performed between the direct pathway coefficients (Fig. 4*B*) and CSE scores (see *Behavioral analysis* and Fig. 1*B*). The average direct pathway coefficient (c'), across all significant indirect paths, was negatively correlated with the magnitude of the CSE (Fig. 4*C*; $r = -0.53$, $P = 0.001$). Because there is a separate c' coefficient estimated for each indirect pathway, it is possible to isolate which mediating paths are accounting for the most

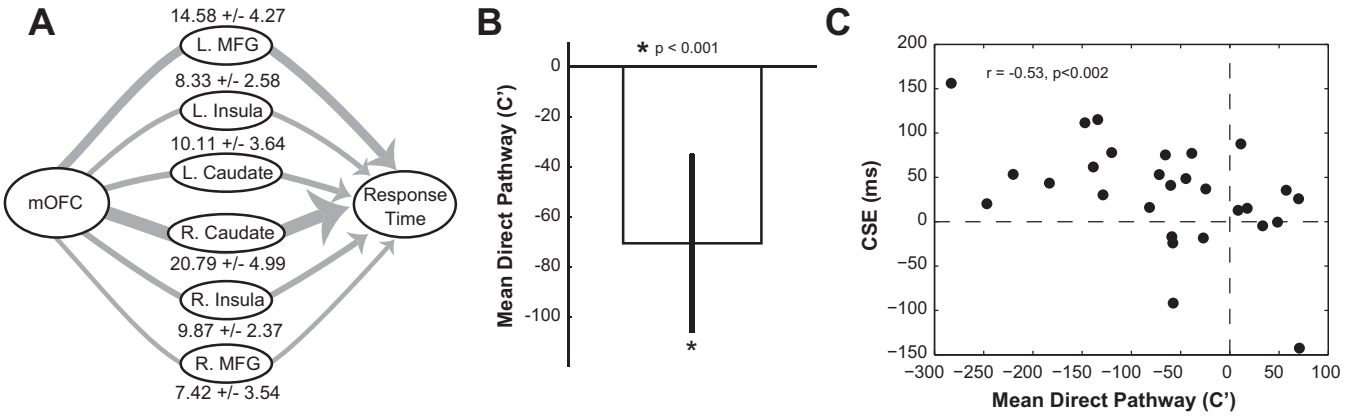

Fig. 4. Mediating network pathways. *A*: statistical mediation analysis revealed 3 bilateral regions that served as indirect pathways mediating the association between mOFC responses and RT on the same trial. Thickness of the lines illustrates the magnitude of the indirect pathway coefficients ($a*b$). Mean indirect pathway coefficient and SD, across subjects, are shown above each mediating ROI. *B*: recovered direct pathway (c') coefficient, showing the residual relationship between mOFC and current-trial RT after accounting for indirect pathways. Plotting conventions as used in Fig. 2. *C*: individual differences analysis showing the association between a subject's direct pathway magnitude and the degree of the congruency sequencing effect (CSE). A Spearman's rank-order correlation coefficient was used in order to reduce the influence of outlier points on the correlation estimation.

variance in the mOFC-RT relationship. After controlling for multiple comparisons ($q < 0.05$), the residual c' coefficients from the bilateral lateral prefrontal (left: $r = -0.55$, $P < 0.001$; right: $r = -0.53$, $P < 0.001$) and bilateral caudate (left: $r = -0.51$, $P < 0.001$; right: $r = -0.54$, $P = 0.001$) pathways were also significantly associated with CSE magnitude. The direction of these relationships implies that the subjects with stronger (i.e., more negative) c' coefficients had the greatest adaptation effect on subsequent trials. More importantly, however, none of the indirect ($a*b$) pathways themselves was associated with the magnitude of CSE (all $P > 0.15$).

These functional imaging results reveal that the mOFC activity correlates with the speed of future responses and is indirectly associated to current-trial RT via mediating pathways in lateral frontal and striatal regions. Most importantly, however, after accounting for these indirect pathways, the strength of the residual mOFC-RT relationship on the current trial predicts the degree of the CSE. Taken together, these findings are all consistent with the hypothesis that the ventral corticostriatal network is associated with trial-by-trial adaptation in cue-conflict tasks.

Convergence of striatal inputs. One common connectivity pattern links all the regions shown in Fig. 4*A*: the orbitofrontal and lateral prefrontal regions all send feedforward projections into the caudate nucleus (Haber and Knutson 2010). Thus, anatomically speaking, the striatum may represent a central integration point during the response selection and adaptation processes.

To evaluate this anatomical hypothesis, fiber tractography data from DSI were used to map out the underlying white matter pathways from 11 frontal regions (see MATERIALS AND METHODS). Figure 5, *A* and *B*, show an example tractography run from a single subject. Streamlines from orbitofrontal, medial wall, lateral frontal, and insula all terminate cleanly within the mask for the caudate nucleus and putamen. Figure 5*B* shows the consistent topography of streamline endpoints, with orbitofrontal fibers primarily terminating ventrally from medial and lateral prefrontal streamlines.

To quantify this topography at the group level, the endpoint location density of streamlines was averaged across subjects for each ROI, and voxels with consistent streamline terminations were determined with a one-sampled *t*-test. Figure 5*C* shows the cortical and subcortical termination fields, across subjects, overlaid on a T2-weighted anatomical template.

Generally there was a gross segmentation of endpoint fields within the striatum that is consistent with previously reported topographies of inputs from frontal areas (Draganski et al. 2008; Haber and Knutson 2010; Verstynen et al. 2012). For example, orbitofrontal projections tended to terminate in the rostral and ventral aspects of the striatum, while lateral prefrontal regions terminated in more dorsal and caudal regions in the body of the caudate. Yet there is also substantial overlap in these endpoint fields along the striatum. This is most clearly seen in the close-up of the striatum in Fig. 5*D*. In general, the streamlines from medial, lateral, and orbital sources shared a moderate degree of overlap in the rostral striatum, particularly the near the shell of the striatum. In fact, the cluster of caudate

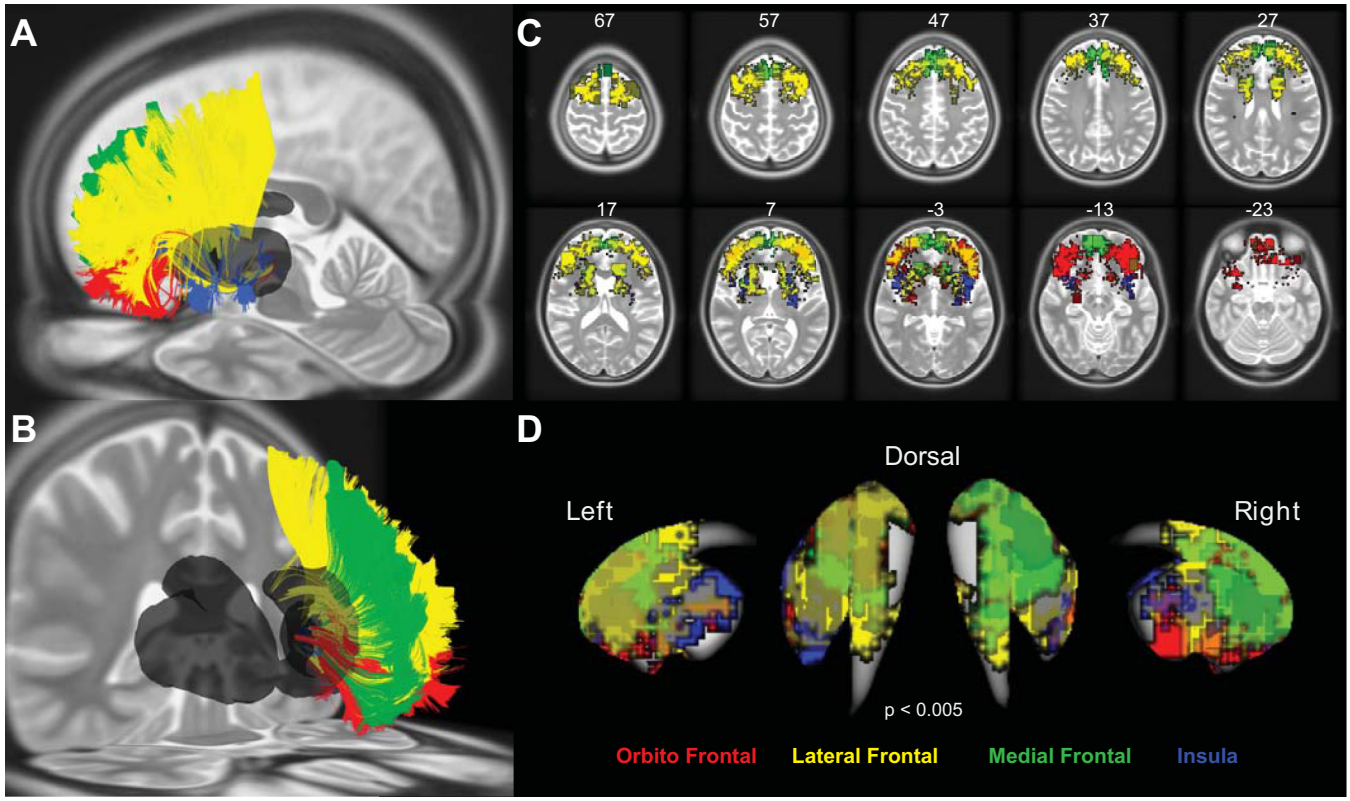

Fig. 5. Topography of corticostriatal projections. *A* and *B*: example deterministic tractography results from a single subject showing tracked left hemisphere projections only. Streamlines are colored based on cortical grouping. *A* shows a lateral view, while *B* shows a medial view. *C*: voxelwise maps showing the locations of highest endpoint density of corticostriatal projections, across subjects. Voxels are thresholded at a $t > 2.75$ and $P < 0.005$, uncorrected. *D*: data shown in *C* projected on a template of the striatal nuclei.

activity observed in the fMRI results (Fig. 2*B*) appears to be situated just between projection fields from all areas tracked, except for the insula, in the rostral caudate. Although given the difference in distortion between functional EPI and diffusion-weighted imaging sequences, it is difficult to attribute the caudate activation to specific inputs from any cortical area.

This overlap of streamline endpoints into the rostral aspects of the caudate is consistent with previous tracer studies in nonhuman primates (Haber et al. 2006). Specifically, Haber and colleagues (2006) propose that integration in the striatum happens when diffuse (i.e., low density) projections from one cortical area overlap with focal (i.e., high density) projection fields from another cortical area. Anatomically, the present results show evidence of these diffuse and focal projection field overlaps. Figure 6*A* shows the mapped white matter projections from the MFG and medial orbitofrontal gyrus for a single subject. The start and end locations of these streamlines are shown in Fig. 6*B*, with the caudate and putamen ROIs shown in black. As is highlighted in the close-up image (Fig. 6*C*), there are several regions of overlap between the two streamline sets.

To better understand the degree of overlap and the density of the projection fields, the endpoint densities along the striatum were calculated for each voxel in the same two pathways. Figure 6*D* shows these densities in three example subjects. *Subject s160* is the same subject as shown in Fig. 6, *A–C* (i.e., showing the same data sets). Consistent with the diffuse overlap hypothesis (Haber and Knutson 2010), the centers of mass for the two projection fields do not overlap; however, there is consistent overlap of the less dense portions of the projection fields in all three subjects.

The degree of overlapping fiber streamlines from adjacent cortical regions to the same striatal voxels (see MATERIALS AND METHODS) was next quantified with an overlap index score (OI; see MATERIALS AND METHODS). The OI defines the percentage of the time that any pair of ROIs overlap on voxels in the mask of the striatal nuclei. This score was averaged across all voxels in the left and right striatum masks separately, providing a composite index for the level of local overlap for each subject and hemisphere. Overall, there was a greater degree of overlap in the right hemisphere (0.40, 95% CI upper bound = 0.50, lower bound = 0.29) than in the left hemisphere (0.22, 95% CI upper bound = 0.29, lower bound = 0.15). It should be noted, however, that even though striatal spiny neurons are known to receive a high concentration of convergent inputs (Kincaid et al. 1998), this score cannot be interpreted as capturing shared collaterals on the same striatal neurons, as this is beyond the spatial resolution of current diffusion imaging methods. Instead, this OI reflects the proximity of inputs from different anatomically defined cortical regions.

The functional significance of this structural measure was assessed with a correlation analysis between the OI and the

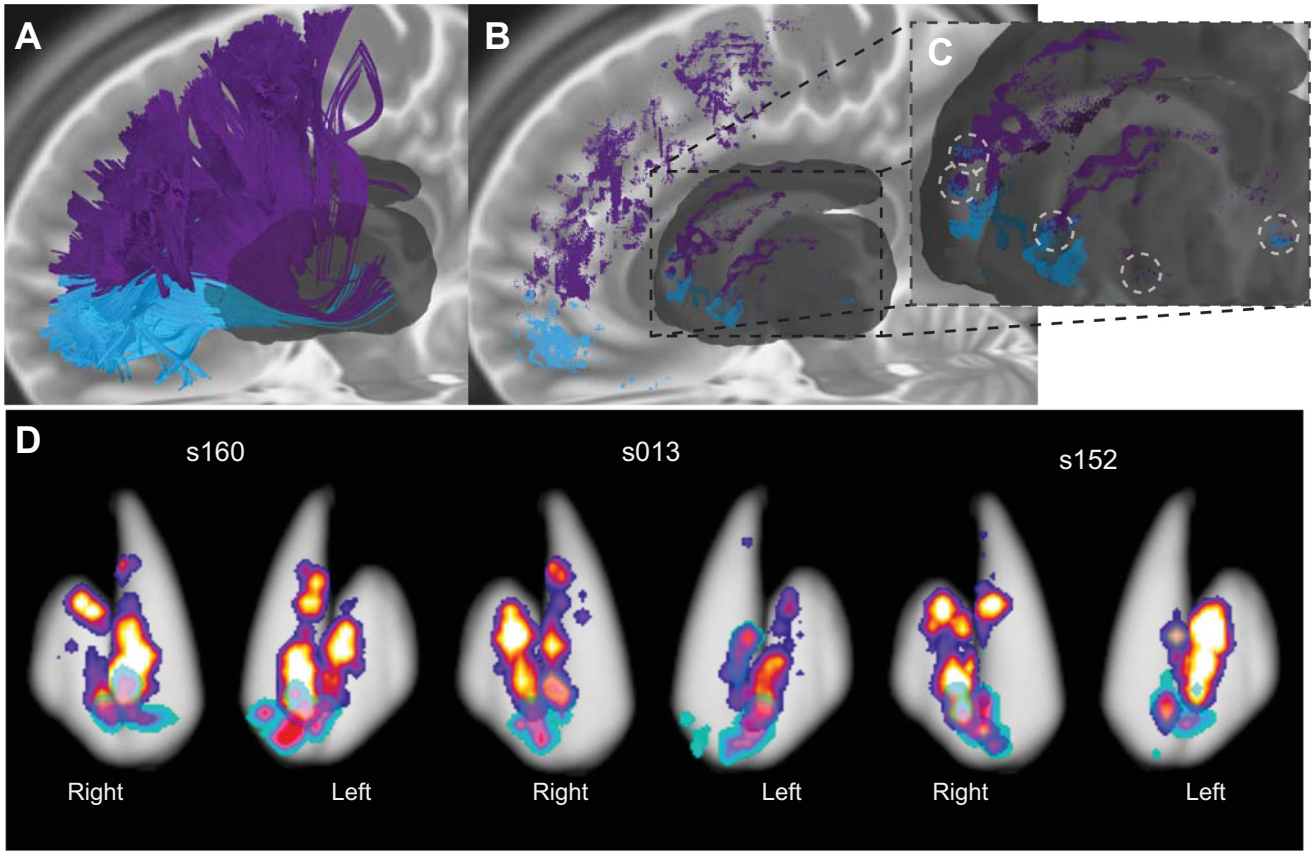

Fig. 6. Overlapping corticostriatal projections. *A*: streamlines from 2 ROIs (middle frontal gyrus, purple; medial orbitofrontal gyrus, cyan) from a different subject from that shown in Fig. 5, *A* and *B*. *B*: streamline endpoint locations for data shown in *A*. Striatum ROI mask shown in dark gray. *C*: close-up of endpoint locations along the striatum. Dashed circles highlight regions where adjacent streamlines overlap. *D*: density maps of streamline endpoints, for the 2 pathways shown in *A*, at each voxel along the striatum for 3 subjects. *Subject s160* is the subject shown in *A*. Warmer maps show endpoint densities from middle frontal gyrus. Cooler maps show endpoint densities from medial orbitofrontal gyrus.

indirect ($a*b$) and direct (c') pathways from the mediation model. While structural overlap did not correlate with the indirect pathways (all $P > 0.39$), the overlap of corticostriatal projections in the left hemisphere was negatively correlated with the magnitude of the direct pathway coefficient (Fig. 7*A*; $r = -0.36$, $P = 0.032$). The direction of this relationship implies that subjects with a greater degree of overlap of anatomical connections into the striatum also had stronger residual mOFC-RT relationships. This pattern was not observed in the right hemisphere (Fig. 7*B*; $r = 0.13$, $P = 0.22$); however, this laterality may be due to the fact that only the left mOFC was included in the calculation of the direct pathway coefficients. Nonetheless, the pathway with the most predictive value for the CSE was itself correlated with local overlap of frontal corticostriatal projections, highlighting a possible structural mechanism for the integration of executive and reinforcement processes during response updating.

A structure-function-behavior pathway. The preceding analysis revealed that subjects with a greater degree of overlapping anatomical connections into the striatum also have stronger, i.e., more negative, conditioned associations between mOFC responses and current-trial RT. Furthermore, stronger conditional mOFC-RT relationships correlated more with the degree of adaptation on future incongruent trials. Taken together, these associations suggest that individual differences in the structural topography of corticostriatal networks may provide a computational constraint on the CSE. In support of this hypothesis, the conditional mOFC-RT effect was found to be a significant indirect pathway linking structural overlap scores and CSE scores ($a*b = 112.4$, 95% CI = 44.0–183.0, $P < 0.001$; Fig. 7*C*). This indirect pathway association was significant despite the fact that the simple correlation between structural overlap scores and the CSE was not significant ($r = 0.15$, $P = 0.23$). Accounting for the indirect associations via the functional network dynamics also did not yield a significant direct path between OI and CSE ($c' = 37.6$, 95% CI = −42.4 to 131.4, $P = 0.19$), suggesting that any relationship white matter had to behavior was fully dependent on indirect associations via functional network processes. Taken together, these patterns of associations suggest that structural overlap of corticostriatal projections has a conditional relationship with behavioral updating through functional network dynamics that bind mOFC activity to RT.

DISCUSSION

Using a standard cue conflict paradigm, the present study revealed that responses in a medial region of the orbitofrontal cortex (on the gyrus rectus) are strongly associated with the speed of upcoming decisions through interactions with frontal and dorsal striatal regions. In particular, stronger (i.e., more negative) evoked responses in the mOFC correlated with faster RTs on the following trial and, after accounting for interactions with lateral prefrontal areas and the caudate nucleus, slower behavioral responses on the current trial. The stronger the relationship between mOFC and current-trial RT, the better a subject adapted to cue congruencies.

The fact that these brain-behavior associations with mOFC were conditioned on concurrent responses in lateral prefrontal cortex and caudate nucleus suggests that any contribution the mOFC has to response adaptation depends on an integration of information from multiple cortical and subcortical sources. One possible mechanism for this integration is via convergent corticostriatal inputs (Averbeck et al. 2014; Haber et al. 2006). Indeed, analysis of structural connections within rostral corticostriatal pathways confirmed that variation in the amount of overlapping projections from adjacent cortical sources into the striatum predicted the efficiency of the functional brain-behavior relationships, i.e., more overlap of white matter projections into the striatum correlated with stronger mOFC-RT associations that, in turn, correlated with stronger behavioral updating.

Perhaps the most striking observation in the present study is the link between mOFC activity and future behavioral responses. So far, the most common type of learning associated with the orbitofrontal cortex is reinforcement learning. Current models of reinforcement learning suggest that the orbitofrontal cortex's role in modifying future responses is to bias correct response mappings based on current trial outcomes, with reward simply influencing the gain of this biasing. Mechanistically it has been proposed that the orbitofrontal cortex helps to modify future behavior by rapidly learning new associations between cues and outcomes through indirect associations with

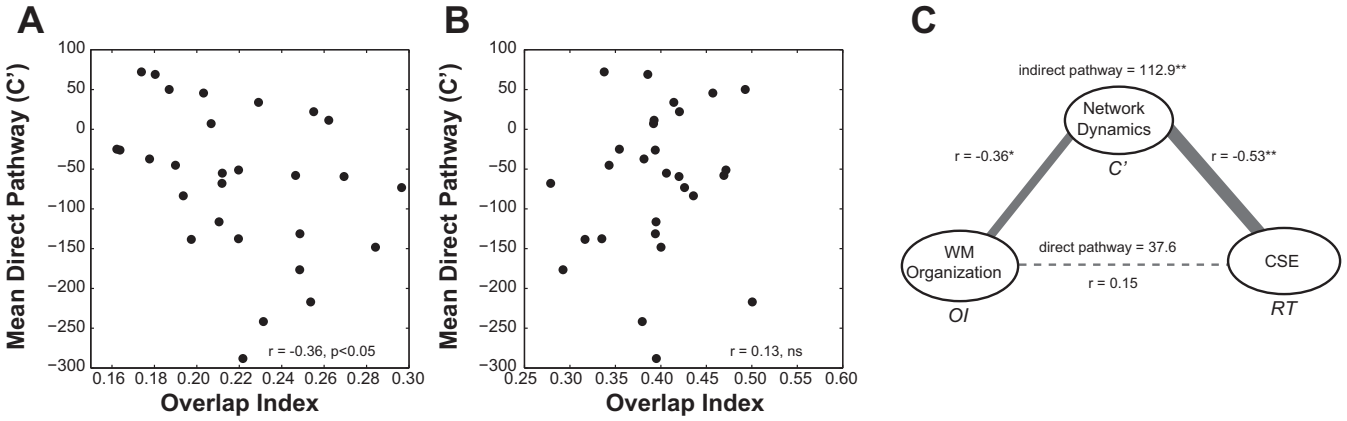

Fig. 7. Structure-function associations. *A* and *B*: Spearman's correlations between direct pathway coefficients from the functional analysis (see Fig. 4) and degree of structural overlap in the fiber tractography analysis for the left hemisphere (*A*) and right hemisphere (*B*) pathways. ns, Not significant. *C*: full structure-function-behavior model showing the associations between structural overlap in the white matter (WM) pathways [overlap index (OI) score], functional brain-behavior links during the Stroop task (c' score from model in Fig. 4, *A* and *B*), and behavioral CSE scores (RT). Data for left hemisphere pathways only. Significant associations: $*P < 0.05$, $**P < 0.002$. *P* values are not corrected for multiple comparisons.

other brain areas (Frank and Claus 2006; Rolls et al. 1996). In this way the orbitofrontal cortex learns to predict outcome expectancies, including but not limited to predicting expected rewards (Schoenbaum et al. 2010). This idea is also consistent with "actor-critic" models of ventral striatal areas during reinforcement learning, in which the striatum contributes to the estimation of internal state values that are compared against incoming sensory information on future trials in order to generate appropriate error signals (O'Doherty et al. 2004). Consistent with this model, ventral striatal neurons thought to be connected to medial prefrontal and orbitofrontal areas have been shown to modulate their firing patterns based on choices made in the previous trial (Kim et al. 2007), suggesting that these neurons retain a memory trace of previous decision outcomes in order to update expected outcomes on the current trial. It remains unclear whether this form of temporal continuity in firing rates is also present in dorsal striatal neurons near the region that was active in the present study. If so, it could provide an electrophysiological mechanism for the BOLD dynamics observed in the present study.

These state-updating models of ventral corticostriatal networks during reinforcement learning are consistent with the pattern of results found in the present study. Greater monitoring of stimulus features that lead to correct responses (i.e., stronger mOFC response) during conflict trials (i.e., when RTs are slower) increases the gain of attending to those features in the immediate future (i.e., faster RTs on the next trial). In this way the orbitofrontal cortex acts as a modulator of executive decisions that are relayed through the prefrontal corticostriatal circuits, rather than a mediator of the CSE itself. This idea is consistent with neural network models of fast reinforcement learning (Frank and Claus 2006; Ratcliff and Frank 2012) in which orbitofrontal projections modulate the sensitivity of go/no-go pathways within the basal ganglia that are triggered by prefrontal selection processes. The observation in the present study that lateral prefrontal and striatal regions served as indirect pathways linking orbitofrontal activity to behavioral responses is consistent with this integration and modulation model.

Mechanistically, orbitofrontal priming of the indirect (go) and direct (no-go) pathways in the basal ganglia requires an integration of information from ventral and dorsal corticostriatal circuits. Although they are traditionally viewed as completely parallel and independent systems (Alexander et al. 1986), it is becoming increasingly apparent that there is a moderate degree of integration across neighboring basal ganglia loops. Concerning frontal cortico-basal ganglia systems, Haber and Knutson (2010) proposed three types of integration pathways that would allow the convergence of executive control and reward information within the basal ganglia: feedback loops relaying information from the ventral striatum to the dorsal striatum via substantia nigra dopamine pathways (Haber et al. 2000), overlap of feedback projections from cortex to the thalamus (which may also interact with "open-loop" afferents from the thalamus to the cortex, allowing for integration of output information from the thalamus; see Joel and Weiner 2000), and local overlap of corticostriatal inputs themselves. This last integration mechanism stems largely from neuroanatomical observations in animals, where diffuse projections from one cortical area terminate in regions of the striatum that contain more dense projections from another cortical origin (Averbeck et al. 2014; Haber et al. 2006; Zheng and Wilson 2002). In humans a similar pattern of diffuse projections was observed in corticostriatal pathways that were tracked with fiber tractography on diffusion-weighted imaging data (Draganski et al. 2008; Verstynen et al. 2012). One of these previous human studies (Verstynen et al. 2012) found an asymmetry in the direction of these diffuse projections, with more streamlines starting in rostral frontal areas and terminating in more caudal striatal regions than vice versa. This means that there were more projections from orbitofrontal areas that terminated in the dorsal and caudal striatum than there were projections from DLPFC that terminated in the rostral and ventral striatum. When considered within the context of the present study, the direction of this asymmetry suggests that this overlap of corticostriatal inputs may be a mechanism of information convergence that is relevant for adapting to congruency sequencing. Indeed, the present study found that individual differences in overlapping corticostriatal projections predicted individual differences in both functional dynamics and brain-behavior relationships during behavioral updating. This strongly suggests that integration of modulatory signals from orbitofrontal cortex and executive control signals from lateral prefrontal cortex happens, at least in part, through common inputs in the striatum. It is entirely possible that reciprocal loops with midbrain areas and convergent feedback projections to the thalamus also predict the CSE; however, visualizing these connections is beyond the capabilities of current white matter visualization methods in humans.

If the medial aspects of the orbitofrontal cortex are modifying future behavioral responses via striatal pathways, then it is reasonable to ask what task-related information the mOFC is representing that is crucial for learning. Often the orbitofrontal cortex is associated with reward processing (see Schoenbaum et al. 2010), yet the present task did not use explicit reward contingencies. An alternative explanation is that the orbitofrontal cortex is not encoding reward per se, but instead encodes the value of a given stimulus. A recent meta-analysis across 206 neuroimaging studies found that medial orbitofrontal areas and rostral striatal regions are often associated with increased activity as the subjective value of a stimulus changes (Bartra et al. 2013). This is consistent with electrophysiological evidence that orbitofrontal neurons modulate their firing rates to changes in subjective value of a stimulus, independent of changes in other stimulus features such as visuospatial characteristics and appropriate motor responses (Padoa-Schioppa and Assad 2009). This value hypothesis of the orbitofrontal cortex implies that the association between mOFC activity and future behavioral responses reflects an updating of stimulus value for a given trial type that fosters optimal performance when presented with similar stimulus classes in the immediate future.

A variant of the value hypothesis that is more closely linked to learning suggests that the orbitofrontal cortex may solve the credit assignment problem between multiple rewards and actions in order to optimize for future decisions (Noonan et al. 2010; Walton et al. 2010). This can be thought of as a multidimensional version of the value hypothesis, whereby every potential response has multiple values associated with it and the goal of the orbitofrontal cortex is to assign the highest weight to the response with the highest value in a given context. Such a hypothesis is consistent with the proposal that the orbitofrontal cortex acts as a state-space monitor for dif-

ferent task and cognitive states during reinforcement learning (Wilson et al. 2014). Consistent with this credit-assignment hypothesis, lesion studies in the macaque have shown that damage to the orbitofrontal cortex impairs an animal's ability to learn to associate different values (i.e., outcome magnitude) with specific actions (Noonan et al. 2010; Walton et al. 2010). Similar lesions in rats have been found to impair an animal's ability to maintain attention to outcome-relevant stimulus parameters, especially for identifying relevant stimulus cues following outcomes that are unexpected (Chase et al. 2012), suggesting that part of the way the orbitofrontal cortex solves the credit assignment problem is through controlled allocation of attention. This attention hypothesis is particularly interesting with regard to the present study, as the mOFC responses observed here may reflect the allocation of selective attentional processes during incongruent trials, which would explain both the selectivity of responses during incongruent trials and the link to RTs on future trials.

It is perhaps curious that the present study did not find an association between activity in medial wall areas, like the ACC, and the CSE (Kim et al. 2013, 2014; Sheth et al. 2011, 2012). As mentioned at the beginning of this article, the behavioral phenomenon of CSE actually reflects an amalgamation of many different cognitive processes, including contingency learning (Schmidt and De Houwer 2011; Ullsperger et al. 2005), feature integration (Hommel 2004; Mayr et al. 2003), conflict adaptation (Botvinick et al. 2001), and, possibly, reinforcement learning (Braem et al. 2012; van Steenbergen et al. 2009). Part of this may be due to the fact that the correlation analysis between BOLD responses and response speed was collapsed across all trial types. It is known that the adaptation of ACC and DLPFC responses during conflict tasks is strictly dependent on the temporal sequencing of trial types (Kim et al. 2013, 2014; Sheth et al. 2012; Weissman and Carp 2013). By searching for a region that predicts all future behavioral responses, this analysis approach was insensitive to those regions whose BOLD response selectively predicts decision speed for specific trial types.

Finally, it is important to point out that within the standard Stroop task it is not possible to isolate the pure CSEs thought to be mediated by conflict monitoring regions like the ACC (Schmidt 2013a, 2013b). This task limitation prevents any elucidation of the underlying role that the orbitofrontal cortex is playing during the CSE itself. Explicitly dissociating the different cognitive components that underlie the CSE, along with their mediating neural systems, will be an important directions for future research to follow.

ACKNOWLEDGMENTS

The author thanks Daniel Weissman and Jean Vettel for their helpful comments on the initial phases of this study, Kirk Erickson for supplying the experimental task used in this study, and Kevin Jarbo, Patrick Buekema, David Creswell, and Fang-Cheng Yeh for their critiques on early versions of the manuscript.

GRANTS

This research was sponsored in part by Pennsylvania Department of Health Formula Award SAP4100062201 and by the Army Research Laboratory under Cooperative Agreement Number W911NF-10-2-0022. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: T.D.V. conception and design of research; T.D.V. performed experiments; T.D.V. analyzed data; T.D.V. interpreted results of experiments; T.D.V. prepared figures; T.D.V. drafted manuscript; T.D.V. edited and revised manuscript; T.D.V. approved final version of manuscript.

REFERENCES

Alexander GE, DeLong MR, Strick PL. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9: 357–381, 1986.

Ashburner J, Friston KJ. Nonlinear spatial normalization using basis functions. *Hum Brain Mapp* 7: 254–266, 1999.

Averbeck BB, Lehman J, Jacobson M, Haber SN. Estimates of projection overlap and zones of convergence within frontal-striatal circuits. *J Neurosci* 34: 9497–9505, 2014.

Banich MT, Milham MP, Atchley R, Cohen NJ, Webb A, Wszalek T, Kramer AF, Liang ZP, Wright A, Shenker J, Magin R. fMRI studies of Stroop tasks reveal unique roles of antreior and posterior brain systems in attentional selection. *J Cogn Neurosci* 12: 988–1000, 2000.

Bartra O, McGuire JT, Kable JW. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76: 412–427, 2013.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. *Psychol Rev* 108: 624–652, 2001.

Braem S, Verguts T, Roggeman C, Notebaert W. Reward modulates adaptations to conflict. *Cognition* 125: 324–332, 2012.

Brittain JS, Watkins KE, Joundi RA, Ray NJ, Holland P, Green AL, Aziz TZ, Jenkinson N. A role for the subthalamic nucleus in response inhibition during conflict. *J Neurosci* 32: 13396–13401, 2012.

Casey BJ, Thomas KM, Welsh TF, Badgaiyan RD, Eccard CH, Jennings JR, Crone EA. Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. *Proc Natl Acad Sci USA* 97: 8728–8733, 2000.

Chase EA, Tait DS, Brown VJ. Lesions of the orbital prefrontal cortex impair the formation of attentional set in rats. *Eur J Neurosci* 36: 2368–2375, 2012.

Chudasama Y, Robbins TW. Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: further evidence for the functional heterogeneity of the rodent frontal cortex. *J Neurosci* 23: 8771–8780, 2003.

Chumbley JR, Friston KJ. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage* 44: 62–70, 2009.

Dayan P, Abbott LF. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (1st ed). Cambridge, MA: MIT Press, 2001.

Diciccio TJ, Efron B. Bootstrap confidence intervals. *Stat Sci* 11: 189–228, 1996.

Diedrichsen J, Shadmehr R. Detecting and adjusting for artifacts in fMRI time series data. *Neuroimage* 27: 624–634, 2005.

Draganski B, Kherif F, Klöppel S, Cook PA, Alexander DC, Parker GJ, Deichmann R, Ashburner J, Frackowiak RS. Evidence for segregated and integrative connectivity patterns in the human basal ganglia. *J Neurosci* 28: 7143–7152, 2008.

Egner T. Congruency sequence effects. *Cogn Affect Behav Neurosci* 7: 380–390, 2007.

Frank MJ, Claus ED. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 113: 300–326, 2006.

Gratton G, Coles MG, Donchin E. Optimizing the use of information: strategic control of activation of responses. *J Exp Psychol* 121: 480–506, 1992.

Grinband J, Savitsky J, Wager TD, Teichert T, Ferrera VP, Hirsch J. The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage* 57: 303–311, 2011.

Haber S, Kunishio K, Mizobuchi M, Lynd-Balta E. The orbital and medial prefrontal circuit through the primate basal ganglia. *J Neurosci* 15: 4851–4867, 1995.

Haber SN, Fudge JL, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci* 20: 2369–2382, 2000.

Haber SN, Kim KS, Mailly P, Calzavara R. Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. *J Neurosci* 26: 8368–8376, 2006.

Haber SN, Knutson B. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35: 4–26, 2010.

Hommel B. Coloring an action: intending to produce color events eliminates the Stroop effect. *Psychol Res* 68: 74–90, 2004.

Joel D, Weiner I. The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience* 96: 451–474, 2000.

Kim C, Chung C, Kim J. Task-dependent response conflict monitoring and cognitive control in anterior cingulate and dorsolateral prefrontal cortices. *Brain Res* 1537: 216–223, 2013.

Kim C, Johnson NF, Gold BT. Conflict adaptation in prefrontal cortex: now you see it, now you don't. *Cortex* 50: 76–85, 2014.

Kim S, Cho YS. Congruency sequence effect without feature integration and contingency learning. *Acta Psychol (Amst)* 149: 60–68, 2014.

Kim YB, Huh N, Lee H, Baeg EH, Lee D, Jung MW. Encoding of action history in the rat ventral striatum. *J Neurophysiol* 98: 3548–3556, 2007.

Kincaid AE, Zheng T, Wilson CJ. Connectivity and convergence of single corticostriatal axons. *J Neurosci* 18: 4722–4731, 1998.

Mackinnon DP, Lockwood CM, Hoffman JM, West SG. A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods* 7: 1–35, 2002.

Macleod CM. Half a century of research on the Stroop effect: an integrative review. *Psychol Bull* 109: 163–203, 1991.

Mayr U, Awh E, Laurey P. Conflict adaptation effects in the absence of executive control. *Nat Neurosci* 6: 450–452, 2003.

Noonan MP, Walton ME, Behrens TE, Sallet J, Buckley MJ, Rushworth MF. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc Natl Acad Sci USA* 107: 20547–20552, 2010.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304: 452–454, 2004.

Padoa-Schioppa C, Assad JA. Neurons in orbitofrontal cortex encode economic value. *Nature* 441: 223–226, 2009.

Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Methods* 40: 879–891, 2008.

Ratcliff R, Frank MJ. Reinforcement-based decision making in corticostriatal circuits : mutual constraints by neurocomputational and diffusion models. *Neural Comput* 24: 1186–1229, 2012.

Rissman J, Gazzaley A, D'Esposito M. Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage* 23: 752–763, 2004.

Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. *Hum Brain Mapp* 31: 798–819, 2010.

Rolls ET, Everitt BJ, Roberts A. The orbitofrontal cortex. *Philos Trans R Soc Lond B Biol Sci* 351: 1433–1444, 1996.

Schmidt JR. Questioning conflict adaptation: proportion congruent and Gratton effects reconsidered. *Psychon Bull Rev* 20: 615–630, 2013a.

Schmidt JR. Temporal learning and list-level proportion congruency: conflict adaptation or learning when to respond? *PLoS One* 8: e82320, 2013b.

Schmidt JR, De Houwer J. Now you see it, now you don't: controlling for contingencies and stimulus repetitions eliminates the Gratton effect. *Acta Psychol (Amst)* 138: 176–186, 2011.

Schmidt JR, Weissman DH. Congruency sequence effects without feature integration or contingency learning confounds. *PLoS One* 9: e102337, 2014.

Schoenbaum G, Roesch MR, Stalnaker TA, Yuji K. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nat Neurosci Rev* 10: 885–892, 2010.

Sheth SA, Abuelem T, Gale JT, Eskandar EN. Basal ganglia neurons dynamically facilitate exploration during associative learning. *J Neurosci* 31:4878–4885, 2011.

Sheth SA, Mian MK, Patel SR, Asaad WF, Williams ZM, Dougherty DD, Bush G, Eskandar EN. Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature* 488: 218–221, 2012.

Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol* XVIII: 643–662, 1935.

Ullsperger M, Bylsma LM, Botvinick MM. The conflict adaptation effect: it's not just priming. *Cogn Affect Behav Neurosci* 5: 467–472, 2005.

Van Steenbergen H, Band GP, Hommel B. Reward counteracts conflict adaptation. Evidence for a role of affect in executive control. *Psychol Sci* 20: 1473–1477, 2009.

Verstynen TD, Badre D, Jarbo K, Schneider W. Microstructural organizational patterns in the human corticostriatal system. *J Neurophysiol* 107: 2984–2995, 2012.

Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* 65: 927–939, 2010.

Watanabe M, Munoz DP. Neural correlates of conflict resolution between automatic and volitional actions by basal ganglia. *Eur J Neurosci* 30: 2165–2176, 2009.

Weissman DH, Carp J. The congruency effect in the posterior medial frontal cortex is more consistent with time on task than with response conflict. *PLoS One* 8: e62405, 2013.

Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81: 267–279, 2014.

Yeh FC, Tseng WY. NTU-90: a high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction. *Neuroimage* 58: 91–99, 2011.

Yeh FC, Verstynen T, Wang Y, Fernandez-Miranda JC, Tseng WY. Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS One* 8: e80713, 2013.

Yeh FC, Wedeen VJ, Tseng WY. Generalized q-sampling imaging. *IEEE Trans Med Imaging* 29: 1626–1635, 2010.

Zheng T, Wilson CJ. Corticostriatal combinatorics: the implications of corticostriatal axonal arborizations. *J Neurophysiol* 87: 1007–1017, 2002.