Supporting Information

Fleming et al. 10.1073/pnas.0910380107

SI Text

Stimuli and Task. Stimuli were presented on a gray background using Cogent 2000 (www.vislab.ucl.ac.uk/cogent.php) running in MATLAB. The "court" consisted of two white tramlines presented either side of fixation, the outer edge of which was viewed at an eccentricity of 12.4° of visual angle (Fig. 1*A* in main text). The "ball" was a filled yellow circle subtending 3.7°. Stimuli were presented using an NEC LT157 LCD projector running at a refresh rate of 60 Hz, viewed by subjects via an adjustable mirror.

Each trial began with a central fixation cross-flanked by two longitudinal white tram lines, presented for a variable interval (750–3,000 ms) in peripheral vision. Participants were asked to maintain fixation and were instructed that not doing so would compromise their performance on the line judgment task. The target ball was presented at either tramline for 66 ms, either overlapping the line (in) or outside the line (out). The difficulty of the decision was manipulated by altering the distance of the stimulus from the outside edge of the tramline. An interstimulus interval of 750 ms followed the offset of the target.

Responses were made using an optical keypad and consisted of a go/no-go decision. Specifically, during each run of trials, participants were required to depress one key with the index finger of their right hand, designated as the "default." Response options (in/out) were presented for 2,000 ms. One of these options was defined as the default by a surrounding black square. Participants continued to depress the default key to select the default option ("accept"); this key was released and a second key pressed to select the alternative ("reject"). On half of trials the target offset was defined as low difficulty and on the other half high difficulty by drawing the offsets from two separate Gaussian distributions defined on the basis of pilot data. The random draw of offsets was further constrained to produce half "out" and half "in" ball positions. The default option was balanced over in/out and over low/high difficulty trials, giving a fully factorial design.

Each participant was given both written and verbal task instructions, before being familiarized with the task format by a short practice block (16 trials). The task involved three runs of 80 trials, with a short break between runs. Participants were informed that they would earn 20p per correct decision and lose 10p for every incorrect decision. This asymmetry in wins and losses was designed to ameliorate the effects of loss aversion on the status quo bias, given previous findings of losses looming approximately twice as large as gains (1). Feedback, in terms of cumulative money earned and lost, was given every 10 trials; trial-by-trial feedback was not given. At the end of the task participants received a bonus payment equivalent to their earnings in their highest-scoring run.

Functional MRI Acquisition and Analysis. Blood oxygen level–dependent (BOLD)-sensitive functional images were acquired using a gradient-echo EPI (echo planar imaging) sequence (48 transverse slices; time to repetition, 2.88 s; time to echo, 60 ms; 3 × 3-mm in-plane resolution; 2-mm slice thickness; 1-mm gap between adjacent slices; z-shim, −0.4 mT/m; negative phase encoding direction; slice tilt, −30°) optimized for detecting changes in ventrolateral and medial frontal cortex (2). Three runs of 188 volumes were collected for each subject, followed by a high resolution (1 mm^3) T1-weighted anatomic scan and local field maps.

The first five volumes of each run were discarded to allow for T1 equilibration. EPI images were realigned and unwarped using field maps (3), and slice-timing correction applied. Each subject's T1 image was segmented into gray matter, white matter, and cerebrospinal fluid, and the segmentation parameters were used to warp the T1 image to the SPM Montreal Neurological Institute (MNI) template. These normalization parameters were then applied to the functional data. For one subject, normalization parameters were estimated from the SPM EPI template because of the unavailability of a T1 image. Finally, the normalized images were spatially smoothed using an isotropic 8-mm full-width half-maximum Gaussian kernel.

Functional MRI (fMRI) time series were regressed onto a composite general linear model (GLM) containing delta (stick) functions representing the onsets of the lines, ball, choice screen, button press (if any), response screen, and cumulative feedback. These delta functions were convolved with the canonical HRF, and low-frequency drifts were excluded with a high-pass filter (128-s cutoff). Short-term temporal autocorrelations were modeled using an AR(1) process. Motion correction regressors estimated from the realignment procedure were entered as covariates of no interest. Statistical significance was assessed using linear compounds of the regressors in the GLM, generating statistical parametric maps (SPM) of *t* values across the brain for each subject and contrast of interest. These contrast images were then entered into a second-level random-effects analysis using a one-sample *t* test against zero.

Small volume correction was applied to a priori regions of interest (ROIs) in the subthalamic nucleus (STN) and medial frontal cortex (MFC). Right and left STN ROIs were defined as 10 × 10 × 10-mm boxes centered on ±10, −15, −5, following Aron and Poldrack (4); the MFC ROI was defined as a 12-mm sphere centered on 0, 27, 30. This volume is representative of coordinates reported in a recent metaanalysis of conflict-related activity in the MFC (5). To quantify the interaction effect, percentage signal change within each STN ROI was extracted for each condition and averaged across subjects and sessions using MarsBar.*

Anatomic Localization of the Interaction Effect. To explore the anatomy of our interaction effect, group-level clusters were projected onto the averaged structural scans from the same subjects in MNI space. With the aid of the atlas of Duvernoy (6), the STN was localized as lying lateral and slightly anterior to the high-signal red nucleus when viewed on an axial slice. On a coronal section, the STN is separated from the gray matter of the thalamus by the zona incerta and the lenticular fasciculus. Using these landmarks, the group maximum for the interaction of decision difficulty and response type (12, −18, 0) was identified as lying ventral to the border of the thalamus, overlapping with the zona incerta/STN. The right-side cluster may extend dorsally into the body of the thalamus (Fig. S2), thus we cannot rule out a contribution of ventral thalamic motor nuclei to the interaction effect. However, the percentage signal changes shown in Fig. 3*B* (main text) and reported in the main text were calculated by averaging over all voxels within a priori STN ROIs, and are thus directly comparable to previous "STN region" activations seen in recent studies investigating response inhibition (4, 7, 8). Further studies are in progress to improve the anatomic localization of the BOLD signal within this region.

Connectivity Analysis. We conducted DCM analysis using SPM8 (www.fil.ion.ucl.ac.uk/spm) to take advantage of the latest software available in this release. DCM models neural dynamics in a system of interacting brain regions by representing the population activity at the neural level with a single state variable for each region (9). The

*Brett M, Anton JL, Valabregue R, Poline JP (2002) Regions of interest analysis using an SPM toolbox. Abstract presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, Sendai, Japan. Toolbox available at http://marsbar.sourceforge.net/.

change in this state vector x in time is modeled as a bilinear differential equation, which for a single input u can be written as:

$$\frac{dx}{dt} = (A + uB)x + Cu$$

In this equation, the A matrix represents the intrinsic (endogenous) connectivity; the B matrix represents the modulatory effect of experimental variables on these connections, and the C matrix represents the driving (exogenous) inputs to the system. To explain regional fMRI responses, DCM combines this model of neural dynamics with a forward hemodynamic model describing how the neural population activity induces changes in BOLD signal (10).

To account for interindividual differences in the precise location of underlying activations, seed voxels for the DCM volumes of interest were individually defined at a liberal threshold of $P < 0.1$, uncorrected. Individual subject peaks for MFC and right inferior frontal cortex (rIFC) were defined as those correlating with the high/low difficulty regressor within 12 mm of the group reaction time (RT) activation maxima (Fig. 4*A* in main text). The MFC maximum was additionally constrained to lie along each individual subject's right cingulate or paracingulate gyri when visualized on the T1 image, and the rIFC maximum was constrained to lie along the inferior frontal gyrus. For one subject, an MFC peak 1 mm outside the search volume, but lying on the right paracingulate gyrus, was accepted (Table S4). The right STN peak was constrained such that it both correlated with the interaction regressor and fell within the STN ROI as defined above. The time series of each area was then extracted as the first eigenvariate of all significant voxels within 4-mm spheres of these maxima. Overall, we were able to extract time series for 14 of 16 subjects: in 1 subject, rIFC and in another, MFC could not be defined owing to a lack of functional activations that met anatomic constraints. These 2 subjects were excluded from the DCM analysis.

We created a new, simplified design matrix for DCM analysis, consisting of three regressors encoding the effects of decision difficulty (*high* = 1, *low* = 0), default acceptance (*accept* = 1, *reject* = 0), and their interaction (*reject_high* = 1, *accept_high* = 0, *reject_low* = 0, *accept_low* = 1), plus motion parameters as regressors of no interest. Because the effect of *accept* was negative in our data, we report this in terms of the symmetric positive effect of *reject* for ease of interpretation. Models were compared at the group level using a random-effects procedure implemented in SPM8 (11). This approach is robust to outliers and uses hierarchical Bayes to optimize the probability density on the models themselves, given data from all subjects. The output of this procedure is an exceedence probability, which describes how likely one model is compared with any other model. Although the class of models in which IFC receives difficulty information was clearly preferred (Fig. S5), the model comparison approach was unable to choose between models 5 and 6 (equivalent exceedence probabilities of 36.9% and 37.1%, respectively). Given that model 6 was more complex, we choose to focus on the most parsimonious model 5 in the main text, reporting here the parameters of the alternative model 6 (Table S5).

Analysis of model 6 parameters supports similar conclusions to that of model 5. We find a modulation of the STN from frontal cortex during default rejection that was greater from rIFC–STN than MFC–STN [paired t test between modulatory parameters in model 6; $t(13) = 2.38$, $P < 0.05$]. This result suggests that the default-related modulation of the STN in the present data is best explained by a direct pathway from rIFC, although we note that DCM cannot rule out indirect routes via unmodeled brain regions. Our intrinsic connectivity results from both models corroborate recent findings using multivariate autoregressive modeling of a similar circuit (12), with significant connectivity between IFC and MFC that is stronger from IFC–MFC, and an absence of baseline connectivity with the STN.

Pre–Supplementary Motor Area Activity in Accept > Baseline and Reject > Baseline. If the same brain regions are involved in both action and inaction they may be cancelled in a contrast that directly compares the two. Evidence for this proposal has been found in both single-unit (13) and imaging (14, 15) studies. This hypothesis would predict additional common regions to be active when comparing these two conditions (*accept* and *reject*) against our implicit task baseline. These contrasts are displayed in Fig. S3, showing consistent activation of the pre-supplementary motor area (pre-SMA) for both *accept* and *reject* decisions. This activation is superior and posterior to the MFC region correlating with RT (Table S2) and corresponds well with the pre-SMA focus isolated in a recent meta-analysis of go/no-go studies (15).

1. Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. *Q J Econ* 106:1039–1061.
2. Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. *Neuroimage* 33:493–504.
3. Andersson JL, Hutton C, Ashburner J, Turner R, Friston K (2001) Modeling geometric deformations in EPI time series. *Neuroimage* 13:903–919.
4. Aron AR, Poldrack RA (2006) Cortical and subcortical contributions to Stop signal response inhibition: Role of the subthalamic nucleus. *J Neurosci* 26:2424–2433.
5. Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004) The role of the medial frontal cortex in cognitive control. *Science* 306:443–447.
6. Duvernoy HM (1999) *The Human Brain* (Springer Wien, New York).
7. Aron AR, Behrens TE, Smith S, Frank MJ, Poldrack RA (2007) Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *J Neurosci* 27:3743–3752.
8. Li CSR, Yan P, Sinha R, Lee TW (2008) Subcortical processes of motor response inhibition during a stop signal task. *Neuroimage* 41:1352–1363.
9. Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *Neuroimage* 19: 1273–1302.
10. Stephan KE, Weiskopf N, Drysdale PM, Robinson PA, Friston KJ (2007) Comparing hemodynamic models with DCM. *Neuroimage* 38:387–401.
11. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. *Neuroimage* 46:1004–1017.
12. Duann JR, Ide JS, Luo X, Li CS (2009) Functional connectivity delineates distinct roles of the inferior frontal cortex and presupplementary motor area in stop signal inhibition. *J Neurosci* 29:10171–10179.
13. Isoda M, Hikosaka O (2007) Switching from automatic to controlled action by monkey medial frontal cortex. *Nat Neurosci* 10:240–248.
14. Mostofsky SH, Simmonds DJ (2008) Response inhibition and response selection: Two sides of the same coin. *J Cogn Neurosci* 20:751–761.
15. Simmonds DJ, Pekar JJ, Mostofsky SH (2008) Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia* 46:224–232.

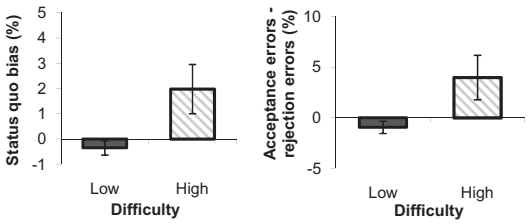

Fig. S1. Replication of the behavioral effects seen in Fig. 2 (main text) in a separate experiment conducted outside of the scanner (n = 18). Two subjects in this cohort also provided data in the main fMRI experiment. The task and experimental protocol were identical to the fMRI design, except the experiment was carried out with the subject seated in front of a computer monitor. The head was stabilized using a chin rest at a distance such that stimulus size and eccentricity was matched to that reported in the main text. Error bars reflect ±SEM.

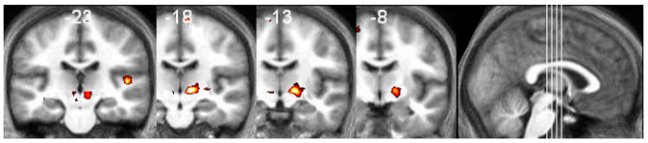

Fig. S2. Multiple coronal views of the interaction effect displayed in Fig. 3*A* (main text).

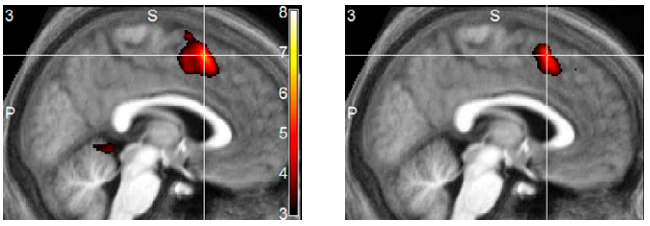

Fig. S3. T-maps for the contrasts *accept* > *baseline* and *reject* > *baseline*, shown in sagittal sections. Midline activation consistent with the location of the pre-SMA (peak voxels; *accept*, 3, 15, 54; *reject*, 3, 15, 54) is seen in both contrasts (both $P < 0.05$, whole-brain corrected; shown at $P < 0.005$, uncorrected).

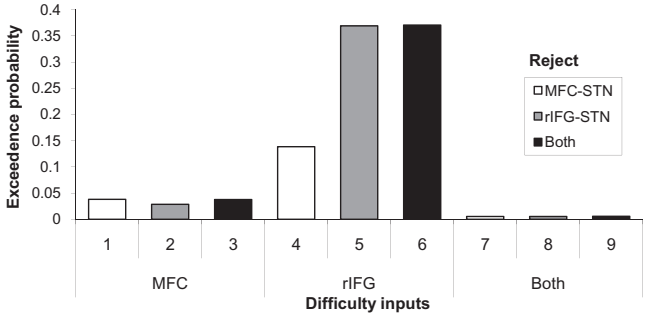

Fig. S4. Exceedence probabilities from the group-level random effects model comparison procedure implemented in SPM8 (1). These probabilities indicate how likely one model is compared with any other model in a given model space, and sum to 1. It can be seen that the models in which difficulty entered into the rIFC were the best fit to the data. A fixed-effects analysis in which a group Bayes factor (GBF) is calculated for each model (2) was also carried out. This analysis gave essentially similar results, with models 4–6 showing roughly equal GBF but differing from the next-best model (model 3) by a GBF of 1.1×10^4.

1. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. *Neuroimage* 46:1004–1017.
2. Penny WD, Stephan KE, Mechelli A, Friston KJ (2004) Comparing dynamic causal models. *NeuroImage* 22:1157–1172.

Table S1. Significant activations in the contrast *reject* > *accept*, defined using a threshold of $P < 0.00001$, uncorrected

Label	MNI coordinate	Voxels (at threshold of 0.00001)	*Z* score	Voxel FWE-corrected *P* value
L postcentral gyrus	−54, −15, 51	575	6.12	<0.001
R cerebellum	24, −51, −30	610	5.95	<0.001
L putamen	−33, 0, −3	430	5.85	<0.001
L precentral gyrus	−57, 3, 30	46	5.79	<0.001
L cingulate gyrus	−6, −24, 48	47	5.49	0.001
R thalamus	15, −15, 3	115	5.45	0.001
L SPL	−24, −54, 51	25	5.35	0.002
R MFG	42, −6, 57	67	5.34	0.002
R IFG	60, 12, 15	23	5.29	0.003
R postcentral gyrus	63, −15, 33	28	5.13	0.008
R postcentral gyrus	36, −33, 57	76	4.93	0.022
R pre-SMA	9, 3, 60	99	4.92	0.023
R cerebellum	30, −81, −21	6	4.89	0.028
R cerebellum	18, −60, −57	25	4.87	0.030
L cerebellum	−36, −45, −30	57	4.85	0.033
L cerebellum	−15, −63, −27	37	4.84	0.035
R posterior temporal lobe	42, −57, 0	3	4.82	0.038
L IFG	−39, 33, 27	6	4.79	0.045

FWE, family-wise error; L, left; R, right; SPL, superior parietal lobule; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; SMA, supplementary motor area.

Table S2. Significant clusters correlating with the RT parametric modulator of the *reject* stick function, defined using a threshold of $P < 0.005$, uncorrected

Label	MNI coordinate	Voxels (at threshold of 0.005)	*Z* score	Cluster FWE-corrected *P* value
R IFC (pars opercularis)	45, 12, 24	140	4.94	0.002
L insula	−39, 9, −12	192	4.32	<0.001
R insula	33, 18, 9	296	4.02	<0.001
L precentral gyrus	−57, 3, 30	46	5.79	<0.001
R MFC (cingulate gyrus)	9, 30, 33	57	3.89	0.005, SVC
L MFC (cingulate gyrus)	−9, 24, 30	23	3.72	0.022, SVC

FWE, family-wise error; R, right; L, left; SVC, small-volume corrected.

Table S3. Significant activations in the contrast *high* > *low* difficulty, defined using a threshold of $P < 0.005$, uncorrected

Label	MNI coordinate	Voxels (at threshold of 0.005)	*Z* score	Cluster FWE-corrected *P* value
L precentral gyrus	−54, 3, 42	824	6.35	<0.001
L/R MFC (paracingulate/pre-SMA)	3, 15, 54	255	3.93	<0.001
L MFG	−24, 0, 63	91	3.81	0.021
R IFC (precentral gyrus/pars opercularis)	63, 6, 27	23	3.76	<0.001, uncorrected

FWE, family-wise error; L, left; R, right; SMA, supplementary motor area; MFG, middle frontal gyrus.

Table S4. Subject-specific peak voxels (MNI coordinates; *x, y, z*) used to define the regional volumes of interest for the DCM analysis

Subject	STN	MFC	rIFC
01	15, −12, −6	9, 21, 39	48, 6, 15
02	12, −12, −9	9, 27, 27	39, 9, 15
03	15, −12, 0	6, 39, 27	51, 6, 18
04	15, −12, −3	3, 27, 42	45, 0, 24
05	15, −15, −6	12, 27, 33	45, 12, 12
06	15, −12, −9	9, 21, 39	57, 12, 24
07	12, −15, 0	9, 27, 39	48, 6, 33
08	N/A	N/A	No voxel
09	6, −18, −9	*12, 18, 42*	51, 3, 27
10	N/A	No voxel	N/A
11	15, −12, 0	9, 21, 36	48, 3, 30
12	12, −12, 0	12, 21, 33	48, 9, 15
13	12, −21, 0	6, 33, 42	54, 12, 21
14	6, −15, 0	12, 21, 39	48, 6, 27
15	15, −12, −6	6, 21, 39	45, 24, 24
16	12, −21, 0	3, 27, 42	45, 24, 24
Mean	13, −14, −3	8, 25, 37 (paracingulate)	48, 9, 22 (pars opercularis)

The MFC coordinate listed in italics for subject 09 fell just outside the 12-mm search volume but lay along the subject-specific paracingulate gyrus, and was thus included in the analysis.

Table S5. Model parameters from winning DCM models 5 (*difficulty* driving rIFC; *reject* modulating rIFC-STN) and 6 (*difficulty* driving rIFC; *reject* modulating both rIFC−STN and MFC−STN)

Model	Matrix	Connection	Parameter (s^{-1})	Group significance (*P*)
5	A (intrinsic)	rIFC–MFC*	0.17*	0.001*
		MFC–rIFC*	0.02*	0.02*
		lFC–STN	0.04	0.34
		MFC–STN	0.01	0.27
		STN–lFC	0.01	0.22
		STN–MFC	0.03	0.22
	B (modulatory)	rIFC–STN*	0.06*	0.03*
		MFC–STN	*N/A*	*N/A*
	C (driving input)	rIFC*	0.03*	0.003*
6	A (intrinsic)	rIFC–MFC*	0.17*	0.001*
		MFC–rIFC*	0.02*	0.02*
		rIFC–STN	0.04	0.34
		MFC–STN	0.01	0.27
		STN–rIFC	0.01	0.22
		STN–MFC	0.03	0.22
	B (modulatory)	rIFC–STN*	0.06*	0.03*
		MFC–STN	0.002	0.10
	C (driving input)	rIFC*	0.03*	0.003*

*Parameters reaching group-level significance.