Hierarchical Bayesian analysis reveals complex neural dynamics of inhibitory control

Research Thesis

Presented in partial fulfillment of the requirements for graduation
with research distinction in Neuroscience in the undergraduate colleges of The Ohio State University

by

M. Fiona Molloy

The Ohio State University
April 2018

Project Advisor: Brandon M. Turner

Contents

Abstract

Cognitive control has been of interest to psychologists and neuroscientists because of its contribution to understanding individual differences, impulsivity, addiction, and obsessive-compulsive disorder. Two tasks used to test cognitive control are the Go/No-Go (GNG) and Stop-Signal (SS) tasks. In the GNG task, subjects are given a cue to respond or withhold a response at the beginning of a trial. The SS task extends this basic paradigm by including the possibility that a "Go" cue may switch to a response-withholding cue. Behavioral and functional magnetic resonance imaging (fMRI) data, extracted for twenty-four regions of interest (ROIs), were collected from eleven subjects who completed both the GNG and SS tasks. In this study, blood oxygenation level-dependent (BOLD) responses were fit using a hierarchical Bayesian analysis to five increasingly complex models of the trial-wise neural activation to improve the signal-to-noise ratio and explore differences in neural activation between response (Go trials) and response inhibition (No-Go/Stop trials). We found that constructing a hierarchy, or adding multiple levels to the model, greatly constrained the predicted BOLD signal by systematically removing outliers. Additionally, increasing model complexity elucidated brain regions that played a role solely in carrying out a response (Go trials). We next replicated these results using the more complicated SS task. We found, from adding a hierarchical structure, that some brain areas showed less activation after a stop signal than during either a Go or No-Go trial. Our results suggest hierarchical modeling is a useful tool in interpreting often noisy fMRI data.

Introduction

Bayesian hierarchical modeling has the potential to be a highly effective tool in understanding neural dynamics by addressing the noisiness of fMRI data and other issues relating to multiple corrections in a systematic way. Hierarchical Bayesian modeling has already been used to improve fMRI research. For example, Bowman, Caffo, Bassett, and Kilts (2008) presented a voxel-based framework for hierarchical Bayesian analysis. Later, Ahn, Krawitz, Kim, Busmeyer, and Brown (2011) showed how hierarchical Bayesian estimates of behavioral model parameters can be used as regressors in fMRI analysis and lead to more constrained results than when the same analysis is run using behavioral parameters estimated by non-hierarchical Bayesian techniques, such as maximum likelihood estimates. Our analysis differs from these two studies in the following ways. First, unlike Bowman et al., the hierarchical model fitting occurs after preprocessing and traditional voxel-based analyses, such as ROI analyses. Thus, our method can be easily integrated into existing pipelines for fMRI analysis. Also, unlike Ahn et al., our instance of hierarchical Bayesian modeling is purely neurally-based and is implemented without making assumptions about behavior.

We tested this framework in the area of cognitive control, and specifically response inhibition. Cognitive control theories, in general, are based on the idea that fronto-parietal connectivity allows for cognitively regulatory abilities (Jung & Haier, 2007; Miller & Cohen, 2001). Additionally, individual differences found in these tasks arise from differences in fronto-parietal connectivity, or whole-brain connectivity to the prefrontal cortex (Cole, Yarkoni, Repovs, Anticevic, & Braver, 2012). Response inhibition is one area of research within the broad domain of cogntive control. Response inhibition research has important applications to individual differences, attention deficit hyperactivity disorder and obsessive-compulsive disorder (Bannon, Gonsalvez, Croft, & Boyce, 2002; Miyake & Friedman, 2012; Schachar & Logan, 1990). Two tasks commonly used to measure response inhibition are the Go/ No-go (GNG) task and the Stop-Signal (SS) task. In the GNG task, subjects are instructed to respond to one stimulus (or set of stimuli), often by invoking a motor response

(i.e. pressing a button), and not to respond to a different stimulus, or set of stimuli. The SS task extends this basic setup by adding a stopping condition, where a Go signal is presented, but after a set delay, a stop signal is presented. One area of research in the response inhibition literature is comparing the neural activation and correlates of GNG and SS tasks (Rubia et al., 2001; Swick, Ashley, & Turken, 2011). Using a hierarchical Bayesian analysis, we are able to gain insight inside the differences between going, not going, and stopping.

We aim to show the benefits of using hierarchical Bayesian modeling in constraining fMRI data and to apply these benefits to understanding the neural dynamics of response inhibition present in GNG and SS tasks. First, we explain the details of the tasks and the methodology for obtaining, preprocessing, and analyzing the fMRI data. Next, we introduce five increasingly complex models of the BOLD response. The results from these model fits are presented in a trifold manner. First, we compare model constraint and predictions. Second, we examine how neural activation changes across response and response inhibition conditions. Third, we evaluate coactivation between brain regions and how differences in coactivation relate to individual differences. We conclude with a summary of the results and a discussion of the limitations and further directions.

Methods

Participants

The eleven participants analyzed in this study were part of another study where multiple tasks, including the GNG, were run in the MRI scanner. These participants later participated a second session for the SS task, so they are included in the current study. All participants were recruited from the Ohio State University and its surrounding community and provided informed consent. The study was approved by the Institutional Review Board of the university. Among the eleven participants (mean age=24.6 years; range from 18 to 48) included in the analysis, there were 5 females and 6 males.

Stimuli

All stimuli were programmed in Matlab using Psychtoolbox extensions (`http://psychtoolbox.org/`) on a Windows PC. The participant lay supine on the scanner bed and viewed the visual stimuli back-projected onto a screen through a mirror attached onto the head coil. In the GNG task, subjects were instructed to press a button when they viewed an A, B, C, D, or E, and to not press any button when they viewed an X, Y, or Z. The SS task contained both of these "Go" and "No go" trials, but also on some trials a Go signal was presented but then after a delay, a Stop signal (square around the letter) appeared on the screen. The GNG task consisted of 75 "Go" and 25 "No-go" trials, for a total of 100 trials. The SS task consisted of 64 "Go" trials, 16 "No-go" trials, and 80 "Stop" trials of 3 different delays (individually fit for each subject, based on response time distributions). There were 160 trials per run, and each subject completed three runs of the SS task, so there were 480 trials total. In this study, our analysis focused on just the first run from both tasks. Figure 1 shows the trial examples for both GNG and SS tasks.

MRI Data Acquisition

MRI recording was performed using a 12-channel head coil in a Siemens 3T Trio Magnetic Resonance Imaging System with TIM, housed in the Center for Cognitive and Behavioral Brain Imaging at the Ohio State University. BOLD functional activations were measured with a T2*-weighted EPI sequence (repetition time = 2000 msec, echo time = 28 msec, flip angle = 72 deg, field of view = 222×222 mm, in-plane resolution = 74×74 pixels or 3×3 mm, and 38 axial slices with 3-mm thickness to cover the entire cerebral cortex and most of the cerebellum). In addition, the anatomical structure of the brain was acquired with the three-dimensional MPRAGE sequence (1×1×1 mm^3 resolution, inversion time = 950 msec, repetition time = 1950 msec, echo time = 4.44 msec, flip angle = 12 deg, matrix size = 256×224, 176 sagittal slices per slab; scan time 7.5 minutes) for each participant.

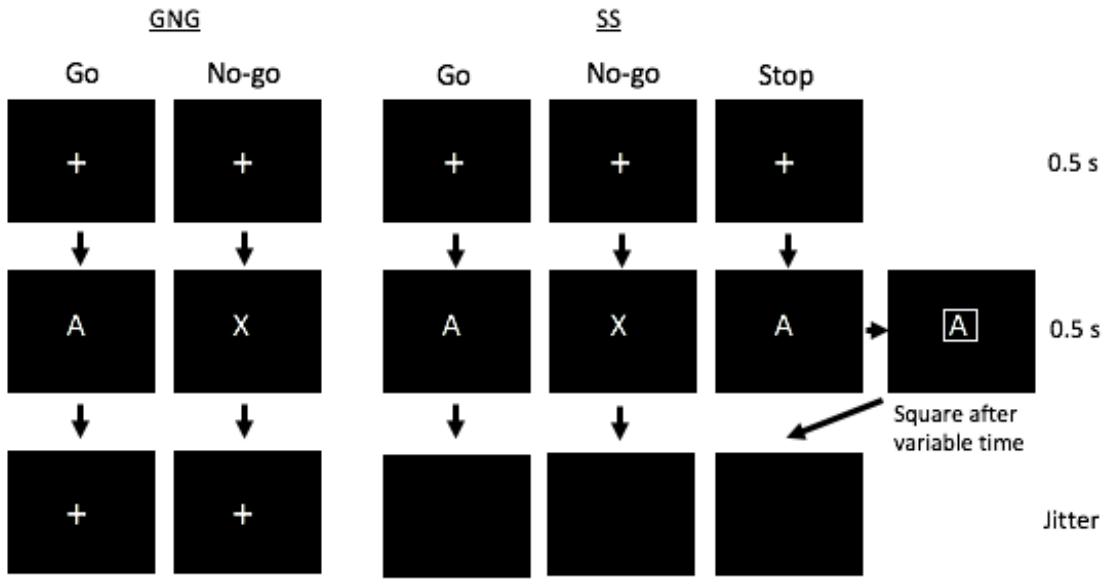

Figure 1. **Example Trials** Diagram showing the example stimulus within a trial. The left panel shows Go/ No-go (GNG) task (one Go trial and one No-go trial), and the right panel shows the Stop-Signal (SS) task (one Go trial, one No-go trial and one Stop trial). For a stop trial, a square around the letter appears after variable time to indicate to inhibit response.

Image preprocessing and analysis

The fMRI preprocessing was carried out using FEAT (FMRI Expert Analysis Tool) in FSL (FMRIB software library, version 5.0.8, `www.fmrib.ox.ac.uk/fsl`). The first six volumes were discarded to allow for T1 equilibrium. The remaining images were then realigned to correct head motion. Data were spatially smoothed using a 6-mm full-width-half maximum Gaussian kernel. The data were filtered in the temporal domain using a non-linear high-pass filter with a 90-s cutoff. A two-step registration procedure was used whereby EPI images were first registered to the MPRAGE structural image, and then into the standard (MNI) space, using affine transformations. Registration from MPRAGE structural image to the standard space was further refined using FNIRT nonlinear registration.

After the neural data was preprocessed, twenty-four regions of interest (ROIs) were extracted. Table 1 shows information about ROIs and their corresponding number labels, used in later figures.

Number	Name	%MNI xyz	nVox (<40)
1.	callosum	[3 -23 29]	208
2.	PCC (posterior cingulate cortex)	[-2 -56 22]	957
3.	preSMA (presupplementary motor area)	[4 21 47]	1952
4.	left angular gyrus	[-44 -72 30]	328
5.	left fusiform gyrus	[-43 -60 -17]	84
6.	left IFG-1 (inferior frontal gyrus 1)	[-37 18 -4]	912
7.	left IFG-2 (inferior frontal gyrus 2)	[-44 9 29]	426
8.	left IPL (left inferior parietal lobe)	[-34 -52 46]	459
9.	left ITG (left inferior temporal gyrus)	[-56 -10 -20]	44
10.	left insula	[-39 -3 7]	41
11.	left MFG (left middle frontal gyrus)	[-3 50 -9]	477
12.	left putamen	[-27 -13 7]	48
13.	left SFG (left superior frontal gyrus)	[-9 57 35]	128
14.	left thalamus	[-6 -16 -2]	72
15.	left ventral striatum	[-1 16 -9]	100
16.	right caudate	[13 10 6]	55
17.	right IFG (right inferior frontal gyrus)	[43 20 12]	2830
18.	right IPL (right inferior parietal lobe)	[48 -44 43]	1400
19.	right MFG (right middle frontal gyrus)	[38 48 -10]	83
20.	right MTG (right middle temporal gyrus)	[49 -66 26]	60
21.	right precuneus	[12 -67 42]	83
22.	right putamen	[31 -11 4]	44
23.	right SFG (right superior frontal gyrus)	[21 49 31]	45
24.	right thalamus	[9 -16 3]	154

Table 1

Regions of Interest *Table showing the number label and full name of each ROI as well as MNI coordinates and number of voxels.*

Model Specification

Five increasingly complex models were constructed to model the neural response during the GNG and SS tasks. Figure 2 shows a graphical diagram of the five models. The complexity of the models increase in a stepwise manner. The first model is the simplest and has no hierarchical component. $N_{i,j}$ represents the observed neural data, where for a given region of interest, there is a blood oxygenation-level dependent (BOLD) response at a time t in response to the presentation of a Go, No-go, or Stop-signal, denoted as stimulus j. The shape of the BOLD response is constructed from convolved hemodynamic response

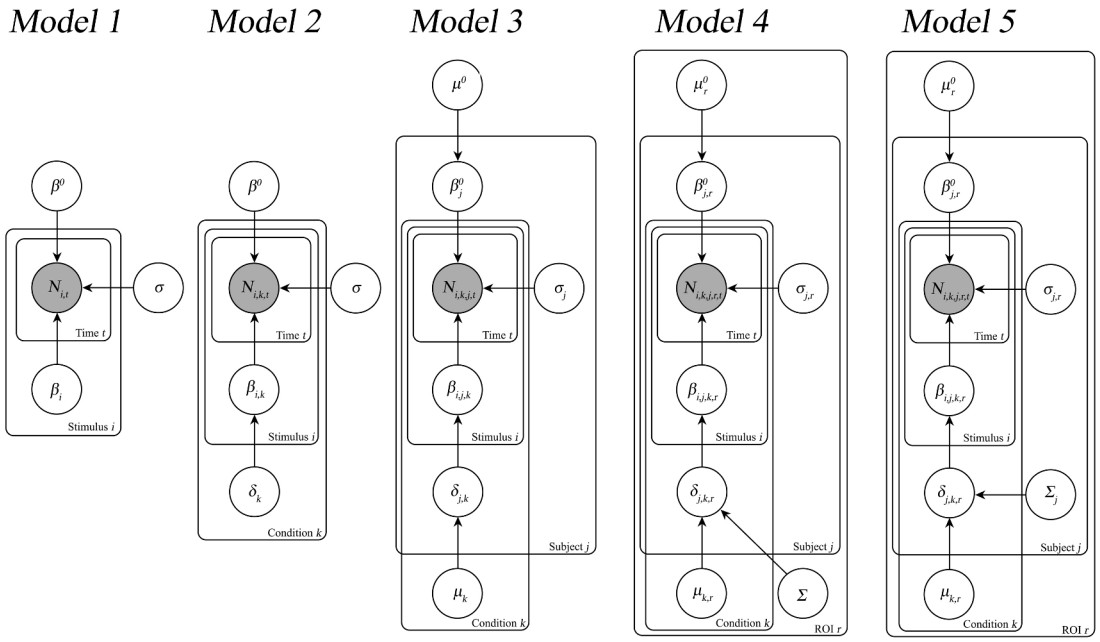

Figure 2. **Model graphical diagrams** Graphical diagrams of the five models. Each node represents a variable in the model, where filled nodes are observed data and white nodes correspond to latent variables. The only filled node pictured in each model is N, the neural data. The design matrix (information about stimuli condition and onset time) were not included in this diagram for visual clarity. Arrows represent relationships between variables and plates represent replications across dimensions (e.g., conditions or subjects).

functions. For a more detailed discussion, see Palestro et al. (2018). The hemodynamic response function (HRF) was chosen to be a canonical form of the double-gamma model implemented in SPM 12 (`http://www.fil.ion.ucl.ac.uk/spm/software/spm12/`):

$$h(t) = \beta h_0(t) = \beta \left(\frac{t^{a_1-1} {b_1}^{a_1} \exp(-b_1 t)}{\Gamma(a_1)} - c \frac{t^{a_2-1} {b_2}^{a_2} \exp(-b_2 t)}{\Gamma(a_2)} \right), \tag{1}$$

where t is time, β is the amplitude of the response, and $\Gamma(x) = (x-1)!$ is the gamma function. The parameters a, b, and c are fixed to their conventional values: $a_1 = 6$, $a_2 = 16$, $b_1 = 1$, $b_2 = 1$, and $c = 1/6$. The only freely estimated parameter is β.

An HRF occurs at every single stimulus presentation, but the shape of the BOLD response is influenced by the timing of HRF following other stimuli presentations and the proximity to other stimuli onsets. When stimuli are close together (e.g. within a 30 second

window), the shape of an BOLD signal for a given stimuli is influenced by the HRFs of other stimuli. The amplitude is increased, as the BOLD response carries over in time. Using data from the design matrix, which contains the onsets of each stimuli (in the case of the GNG task, when a Go or No-go letter appears on the screen), we can predict when a HRF will occur. fMRI measures are delayed from when they actually occurred, but the BOLD response is linearly time invariant (LTI), which means that the BOLD response is delayed the same amount of time as the neural activation. This allows us to assume that the HRF starts at point t when the stimulus was presented. Also, the amplitude, β, can be observed relatively across subjects, conditions and ROIs. This is because β is linearly related to the strength of neural activation in a given area.

The process of this shifting and change in amplification is called convolution. The first step, shifting, can be represented by the following equation:

$$\begin{aligned}(f * h)(t) &= \int_{-\infty}^{\infty} f(\tau)h(t-\tau)d\tau \\ &= \int_{-\infty}^{\infty} h(\tau)f(t-\tau)d\tau \quad \text{(commutativity)}. \end{aligned} \tag{2}$$

where $h(t)$ is the canonical HRF from Equation 1 and $f(t)$ is a boxcar function that contains the timing of stimulus presentation. f(t) contains the t of a stimulus' onset, and zeros at every other time point. This centers the HRF at t for when a stimuli was presented. The next step of convolution addresses the amplitude change that occurs when HRFs are close together. To integrate each individual HRF we calculate from above, we use beta-series regression (Mumford, Turner, Ashby, & Poldrack, 2012; Rissman, Gazzaley, & D'Esposito, 2004), which sets individual regressors for each trial to use in a generalized linear model (GLM).

The neural likelihood can then be defined as the baseline activation (β_0) plus the sum of the convolved and amplified HRFs with an added error term:

$$\begin{aligned}\mathbf{N}(t) &= \beta_0 + \sum_{i=1}^{R} h_i(t) + \epsilon(t) \\ &= \beta_0 + \sum_{i=1}^{R} \beta_i h_{0,i}(t) + \epsilon(t), \end{aligned} \tag{3}$$

R is the number of stimulus presentations, in the GNG task R=100 (one stimulus per trial) and in the SS task R=240 (one per Go/ No-go trial and two per Stop trial, one per each stimulus shown). Again, the only free parameters are the amplitudes of β. The distribution of the added error term ϵ(t) is assumed to be normal, centered at 0, with a standard deviation of σ, which is freely estimated:

$$\epsilon(t) \sim \mathcal{N}(0, \sigma).$$

The distribution of the neural data is also normal, with the mean as the design matrix, $\mathbf{X}$, multiplied by single-trial βs and a standard deviaton of σ:

$$\mathbf{N} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma), \tag{4}$$

All of the above is specifically for Model 1, but also provides the basis for all five models. The second model adds a hierarchical structure across conditions. For the Go/ No-go task there are 2 conditions: Go and No-go. For the stop-signal task, there were four conditions: Go, No-go, stop-signal presented before a response was made, and a nuisance regressor, for when a stop-signal was presented after a response was made. In this model, an additional free parameter is added, the hyperparameter on β, δ. Each condition, k, has a corresponding δ parameter. Model 3 constructs an additional level of hierarchy across subjects. Two more freely estimated are added with this model. The first, μ^0, is a hyperparameter on the each subjects baseline activation β_j^0. The second added free parameter is μ_j, a hyperparameter of $\delta_{j,k}$.

Models 4 and 5 add a hierarchical structure across ROIs. These models add one free

parameter, Σ. Σ informs the prior for δ, which in Model 4 is:

$$\delta_{j,k,r} \sim \mathcal{N}_r(\mu_k, \Sigma),$$

and in Model 5 is:

$$\delta_{j,k,r} \sim \mathcal{N}_r(\mu_k, \Sigma_j),$$

where the notation is consistent with Figure 2, with subject j, condition k, and ROI r. The difference between Model 4 and Model 5 is that the ROI covariance matrix for Model 4 is collapsed across subjects (Σ), whereas Model 5 has one ROI covariance matrix for each individual (Σ_j). The priors for μ_k and Σ are distinctly specified for Models 4 and 5. Note that the parameters for the priors on μ_k and Σ can be set arbitrarily. For Models 4 and 5, the prior for μ_k is normally distributed

$$\mu_k \sim \mathcal{N}_p\left(\phi_0, s_0\right),$$

We set ϕ_0 to be a vector of 24 zeros and s_0 to be a twenty-four by twenty-four matrix of zeros with the diagonal set to 1. For model 4, the prior for Σ is an inverse Wishart distribution,

$$\Sigma \sim \mathcal{W}^{-1}(I_0, n_0),$$

where I_0 is another twenty-four by twenty-four matrix of zeros with the diagonal set to 1 and n_0 is 24. For Model 5, the prior is the same, but there is a separate one for each subject j.

Fitting details

To fit the models, we used Just Another Gibbs Sampler (JAGS; Plummer (2003)). All of the models across the two tasks had three chains, but took on one of two combinations of adaptation, burn-in, and sampling iterations. The first, longer procedure, was used for Models 1, 2 and 3 in the GNG task and Model 1 in the SS task. In this procedure, model

initialization ran for 2,000 adaptations. After initialization, 4,000 samples were discarded as burn-in. Then, the posterior sampling ran for 6,000 iterations. Thus, with three chains, there was a total of 18,000 samples for each parameter.

Because of computational complications, the sampling lengths was shortened for Models 4 and 5 of the GNG task and Models 2 through 5 of the SS task. In this procedure, model initialization ran for 1,000 adaptations, followed by a burn-in period of 2,000 iterations. The posterior sampling then ran for 3,000 iterations. With the three chains again, there were a total of 9,000 samples for each parameter.

For all models, the chains were plotted and visually checked for convergence. All of the hyperparameters (when applicable) and σ and a sample of the predicted neural response (N) and single-stimuli amplitudes (βs) were checked for each subject and ROI.

Model Constraint

The first step in our analysis was to compare model constraint. We compared model fits in three stages. We first looked at the models' predictions of neural responses across time. All five models predicted the neural responses across the time series and therefore can all be directly compared. Additionally, because we observed this neural response, we can see how well the predictions map onto the observed data. Next, we compared model predictions of β across the five models. Unlike in the first comparison, β is latent and thus cannot be examined in relation to any real data. Lastly, we compared model predictions of δ. Because δ is the hyperparameter of β, it is only present in the models with a hierarchical component. Thus, only models 2 through 5 were analyzed. We separated these comparisons by task, starting with the GNG task.

Go/ No-go

Time Series Predictions. All five of the models predicted neural activation at every point of the time series. Figure 3 compares the abilities of the first 3 models in predicting the neural activity of the left insula (ROI 10) across the time series for the GNG

task. Each column corresponds to a model and each row corresponds to a subject. The first three subjects were chosen for illustrative purposes, but represent the general trend shown across subjects. The black dots in each subplot represent the real, observed BOLD response. The solid red line is the mean of the posterior of the predicted neural data across the time series and the dotted red lines represent the 95% predictive interval. The neural predictions from Models 4 and 5 did not visually differ from the predictions from Model 3, and are not included in this figure.

In this figure, Model 1 outperforms Models 2 and 3 in capturing subjects' BOLD response in the left insula. This trend was also observed across subjects and ROIs. Model 1 has a lot of variability and closely captures the observed neural data. The 95% predictive interval continues to closely capture the observed neural data, containing essentially all of the observed data points. Despite their increased complexity, Models 2 through 5 have much less variance and their 95% predictive intervals include less of the observed data than Model 1. Importantly, these results do not necessarily mean that Model 1 best captures all of the data. The next comparison focuses on the constraint of β estimates.

Constraint on Beta Estimates. While increasing model complexity did not improve time series predictions, it did greatly improve single-trial beta estimates (β). Figure 4 compares predictions of β from the first three models, again with left insula during the GNG task. Similar to Figure 3, the rows correspond to the first 3 subjects and the columns correspond to the first 3 models. Unlike in Figure 3, however, there is no real data for comparison because β is latent, or unobserved. The dotted lines refer to the range of the posterior estimates and the red boxes denote the interquartile range.

This figure shows a representation of the large shrinkage effect on model estimations of single trial β, especially when going from a nonhierarchical model (Model 1) to a hierarchical model (Model 2). Additionally, model hierarchy reintroduced some variability in the model estimates. Models 4 and 5 did not drastically improve upon the estimates of Model 3, so they are not pictured. Importantly though, Models 4 and 5 present more information regarding correlations between ROIs, as well as more constraint on the posteriors of other

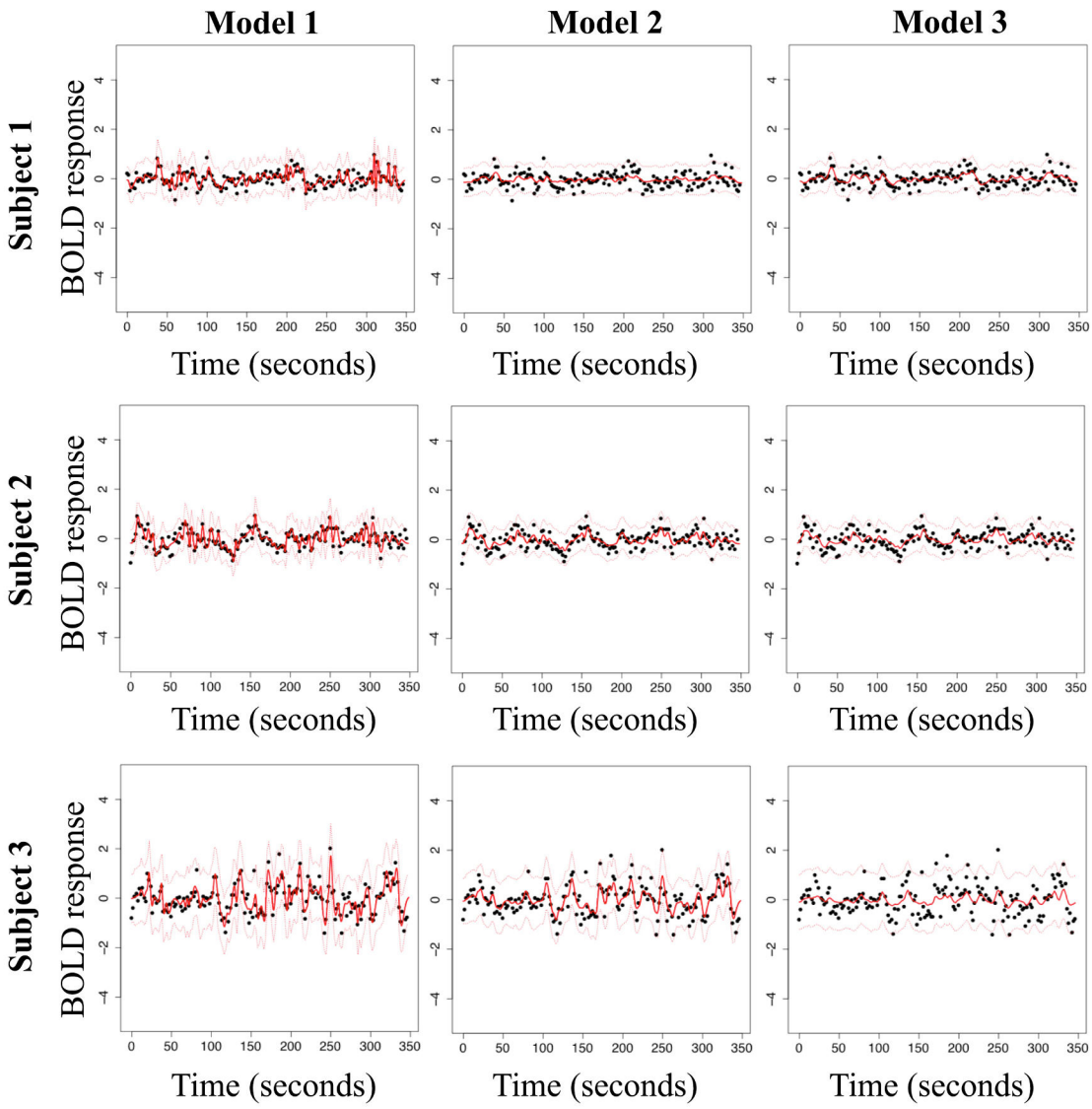

Figure 3. **Time Series Fits** Model predictions of neural activity of the left insula in the GNG task. Rows correspond to subjects and columns correspond to model. The black dots in each subplot are the real, observed BOLD response. The solid red line is the mean of the posterior predicted neural data across the time series and the dotted red lines represent the 95% predictive interval.

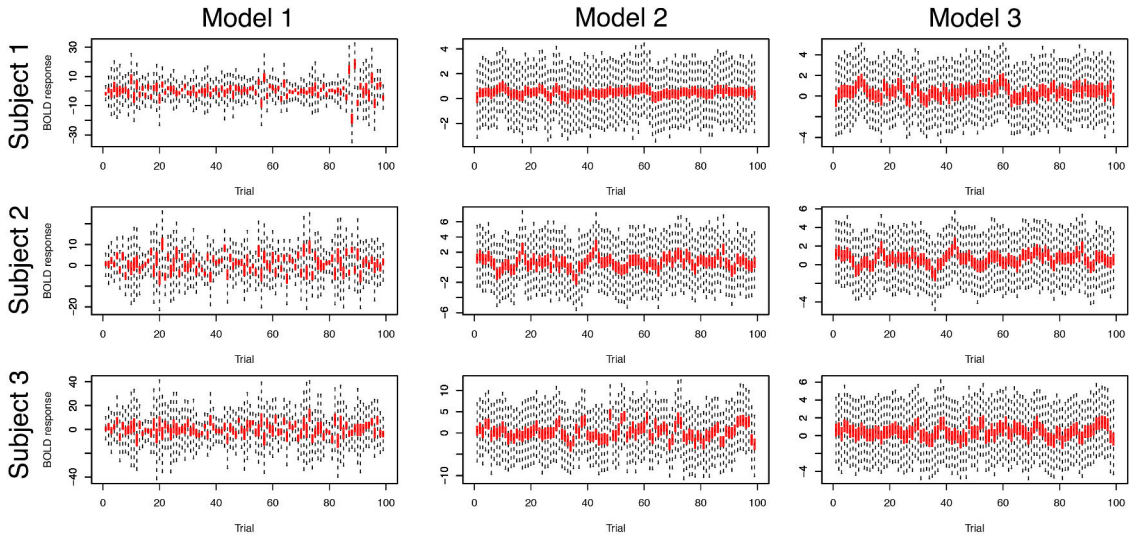

Figure 4. **Constraint on Beta Estimates in GNG** Representative plots of the constraint introduced when constructing a hierarchical component into a model. These plots are for model fits to the GNG task. These single trial estimates are all for ROI 10, the left insula, an area implicated in both the GNG and SS tasks. Each row corresponds to a different subject and each column corresponds to a model. The red box shows the interquartile range and the dotted lines show the range of the posterior.

parameters, such as the β hyperparameters (δ).

An example of the constraining effect of model complexity on δ is shown in Figure 5. This figure shows the joint distribution of δ_{Go} and δ_{No-Go} for the left insula in Model 3 (left panel) and Model 4 (right panel). In this figure, each red point corresponds to a different subject. The x-coordinate of the point is the mean of the δ posterior for the Go condition and the y-coordinate is the mean of the δ posterior for the No-Go condition. The error bars are two standard deviations away from either the δ_{Go} mean if horizontal or the δ_{No-Go} mean if vertical. The diagonal dotted line represents indifference. In other words, if a point lies on that line, a subject has an equal predicted BOLD response in trials when they are told to press a button and trials when they are told not to press a button for that specified ROI.

While single-trial β estimates look essentially identical for Model 3 and Model 5, the comparisons of δ_{No-go} versus δ_{Go} in the two models, shown in this figure, highlight the

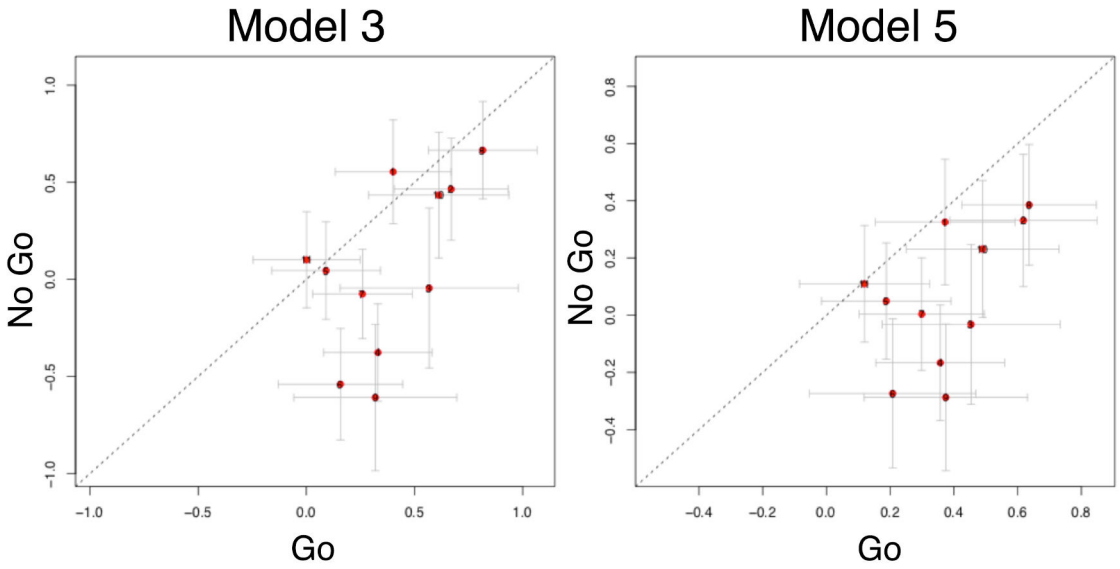

Figure 5. **Constraint on Beta Hyperparameters in GNG** This plot shows the joint distribution of δ_{Go} and δ_{No-Go} for the left insula in Model 3 (left) and Model 5 (right). Each red point corresponds to a different subject. The x coordinate of the point is the mean of the δ posterior for the Go condition and the y coordinate is the mean of the δ posterior for the No-Go condition. The error bars are 2 standard deviations away from either the δ_{Go} mean if horizontal or the δ_{No-Go} mean if vertical.

constraint offered by Model 5. In Model 5, the mean of the posterior of δ_{Go} is greater than δ_{No-go} in all eleven subjects. If we looked only at Model 3, the result is less robust. For example, in Subjects 1 and 11, the mean of δ_{No-go} is actually greater than the mean of δ_{Go}.

Stop-signal

As previously stated, the models for the SS task are more complex as there are two additional conditions. To see how this added complexity would affect model predictions, we again reviewed single-trial (or in this case, single-stimulus) β estimates and the hyper-parameters for these estimates. The time series predictions were also observed, but are not pictured. The time series analysis in the SS task yielded the same results as the analysis in the GNG task, namely that Model 1 fit the observed neural data more closely than the other models.

Constraint on Beta Estimates. Figure 6 replicates the finding from the GNG task that adding even one layer of hierarchy greatly constrains β estimates. The layout and

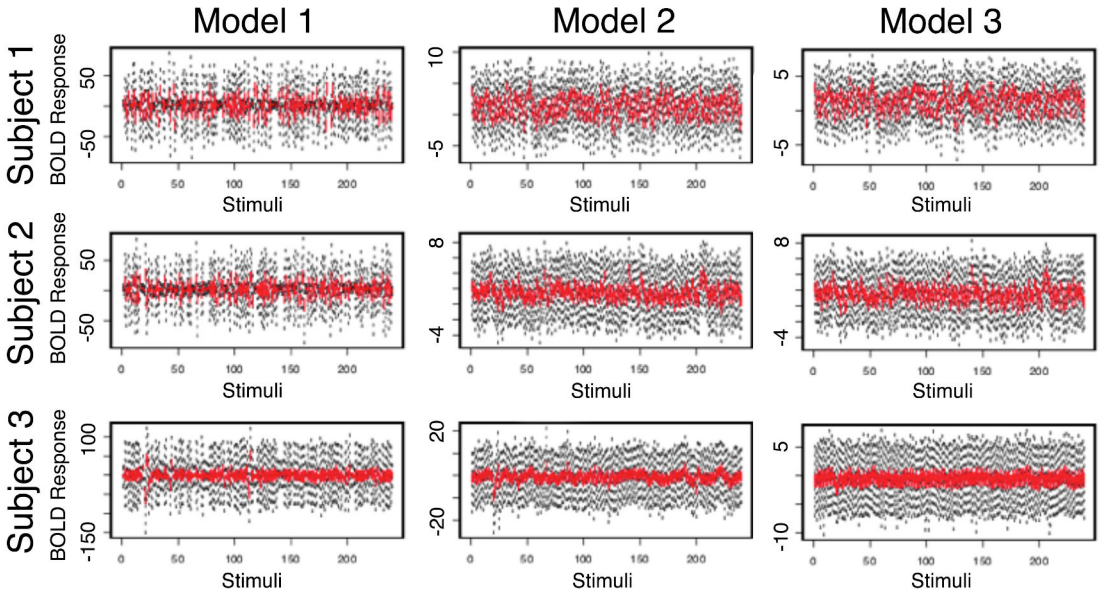

Figure 6. **Constraint on Beta Estimates in SS** Representative plots of the constraint introduced when constructing a hierarchical component into a model. These plots are for model fits to the SS task. These single-stimuli estimates are all for ROI 5, the left fusiform gyrus. Each row corresponds to a different subject and each column corresponds to a model. The red box shows the interquartile range and the dotted lines show the range of the posterior.

notation of this figure are identical to Figure 4, but in this case, the β estimates are for the left fusiform gyrus in the SS task. Additionally, because this is the SS task, the estimates are not for each trial, like in the GNG task. In both tasks, the models estimate a β for the onset of each stimulus (i.e. the letter "A" signifying "Go"), but in the GNG task this is equivalent to trial. In the SS task, however, in trials with a stop-signal, there are two stimuli presented in a single trial (a "Go" cue followed by a "Stop" cue), and thus for some trials there are two β estimates.

Perhaps the largest effect of introducing a hierarchical structure in Figure 6 is the narrowing of the posteriors. This effect can be seen in the changes of the scale of the y-axes from one model to another. The most dramatic change, both in the figure and across all subjects and ROIs, occurs from the transition from Model 1 to Model 2. This effect was also seen in the GNG task, represented by Figure 4, but the posteriors are even wider for the SS estimates. For some subjects/ROIs, the range of posteriors is over 200. An example

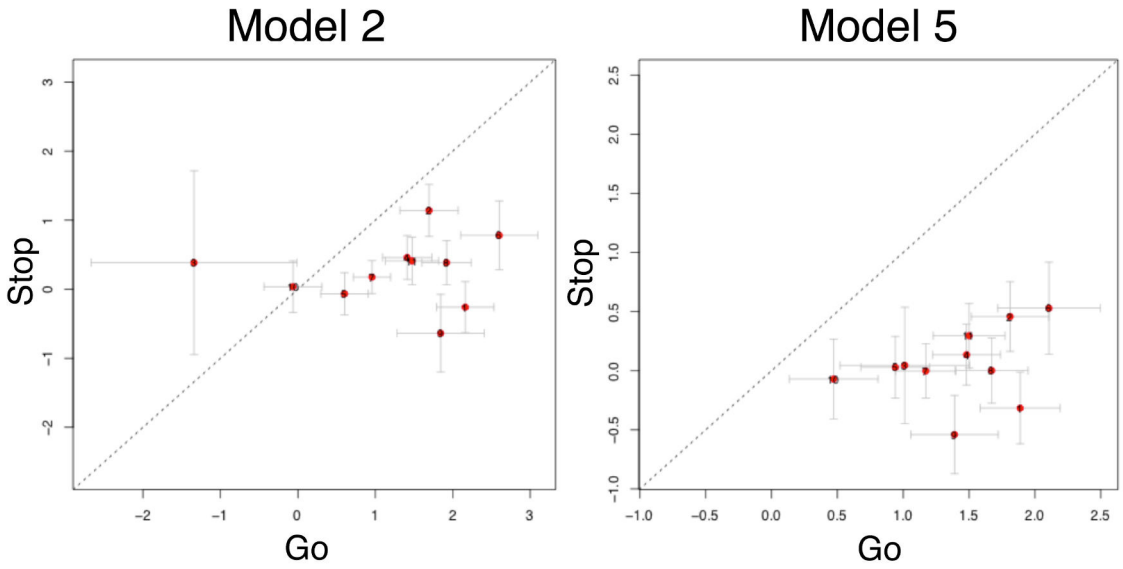

Figure 7. **Constraint on Beta Hyperparameters in SS** This plot shows the joint distribution of δ_{Go} and δ_{Stop} for the left fusiform in Model 2 (left) and Model 5 (right). Each red point corresponds to a different subject. The x coordinate of the point is the mean of the δ posterior for the Go condition and the y coordinate is the mean of the δ posterior for the Stop condition. The error bars are 2 standard deviations away from either the δ_{Go} mean if horizontal or the δ_{Stop} mean if vertical.

of this can be seen in Subject 3's predictions of the left insula's BOLD response for each stimulus, with a range of -150 to 100. Model 2 scales this down to -20 to 20, and Model 3 constrains it further.

In the GNG task, we saw that the most complex model led to different and more constrained estimates of δ than less complex models. Figure 7 replicates this result in the SS task. This figure is set up identically to Figure 5 with a few exceptions. First, δ_{Go} (x-axis) is compared with δ_{Stop} (y-axis) instead of δ_{No-go}. Importantly, this stop condition includes only the stop signals that appeared before a response condition was made. Additionally, these δ predictions are for the left fusiform gyrus (not the left insula like in Figure 5), consistent with the ROI used for the time series analysis in Figure 6. Lastly, this figure compares Models 2 (left) and 5 (right).

In Model 5, all eleven subjects show more activation in the Go condition than in the Stop condition. In Model 2, however, only nine subjects show this trend in activation.

Additionally, in the Model 2 subplot, Subject 3 shows a greater activation for δ_{Stop} than δ_{Go}, with a relatively larger difference (or distance from the line of indifference) than observed in other subjects. However, in the Model 5 subplot, Subject 3 shows the opposite, with more activation in δ_{Go} than in δ_{Stop} and is closer to the group mean. Furthermore even with $\pm$ 2 standard deviations from the mean, each subject's δ_{Go} is larger than δ_{Stop}. In conclusion, increasing model complexity in the SS task adds more constraint in both the β and the δ estimates.

Differences in Activation Across Tasks and Conditions

To explore activation differences across task, we looked at the β hyperparameters, δ. δ defines the posterior from which the single-trial βs are sampled. For each model, in both the GNG and SS tasks, each condition has its own δ value, and thus provides a consistent way to compare how the change in BOLD activation differs across subjects. In the GNG task, there are 2 conditions, and thus 2 δs, δ_{Go} and δ_{No-Go}, in Models 2 through 5. We had four different δ distributions for this task, Go and No-go, similar to the GNG task, but also two stop δs. The first is when a subject did not respond before the onset of a stop-signal, and the second is the nuisance regressor, where a subject responded before a stop-signal was presented, but still observed the stop signal. We will report the δs only from M5, because as shown from the previous section, M5 provides the most constraint, most notably in these δ values (Figures 5 and 7).

Go/ No-go

The aggregated group results of the δ values by condition in the GNG task are shown in Figure 8. Each numbered dot corresponds to an ROI. The location of the dot is approximated for visual clarity. The color of the dot represents the percent change in BOLD signal. Cooler colors represent a smaller, or more negative, percent change in BOLD, and warmer colors represent a larger percent change in BOLD. We took the average $\delta_{condition}$ for each subject across all iterations and chains and then took the group average. We can see areas

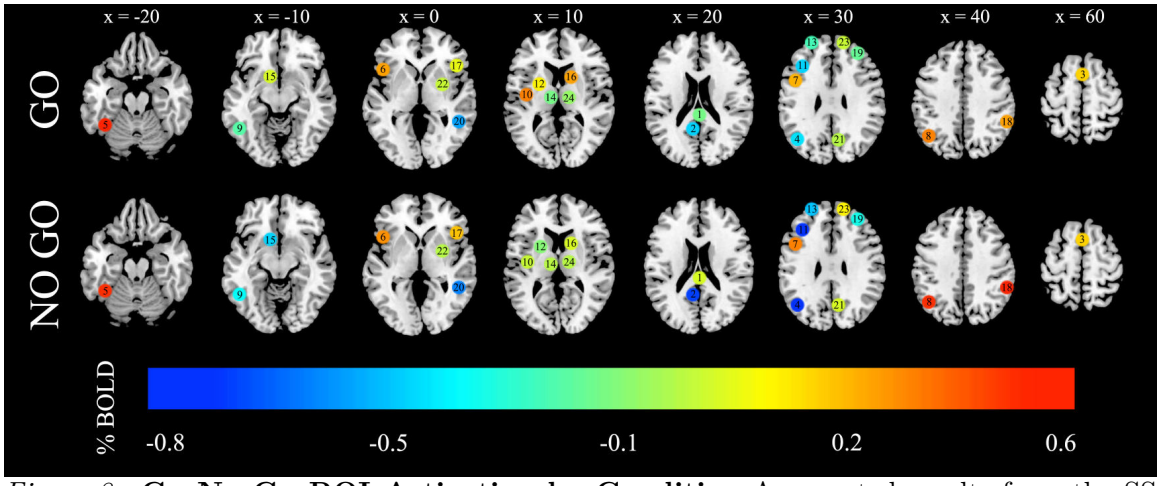

Figure 8. **Go No Go ROI Activation by Condition** Aggregated results from the SS task across individuals for the mean of the δ distribution (hyperparameter on the single-trial β) for each ROI. Cooler colors represent a smaller or more negative percent change in BOLD signal, i.e. less activation, and warmer colors represent a larger percent change in BOLD signal, i.e. more activation. The rows correspond to the mean of δ for each condition, with Go on the top and No-go on the bottom.

that are approximately equal, such as the left fusiform gyrus (ROI 5, column: $x = -20$), right putamen (ROI 22, $x = 0$), and right thalamus (ROI 24, $x = 10$). Other areas show higher activation in the go condition, such as left ventral striatum (ROI 15, $x = -10$), left insula (ROI 10, $x = 10$), and right caudate (ROI 16, $x = 10$). Lastly, some areas show higher activation in the No-go condition, such as left IPL (ROI 8, $x = 40$) and right IPL (ROI 18, $x = 40$).

While this figure can give an overview of the results, it is not possible to appreciate the individual differences known to be important in this task. Additionally, one subject could be skewed in one direction, but this would not be representative of the actual difference in task. Therefore, we also looked at the joint distribution of the delta parameters on an individual level. When discussing these results, we will use the phrase "more activation," which in this case means that the mean of posterior for δ_k for one condition was greater than the mean of the posterior of the δ_k for a different condition. Four ROIs showed more activation for one condition over the other in nearly every subject. The left insula (ROI 10), left putamen (ROI 12), and left ventral striatum (ROI 15) showed more activation in the

Go parameter than in the No-go parameter. This trend was present in all eleven subjects. There was no ROI that showed more activation for No-go over Go in all eleven subjects. However, one ROI, the right inferior parietal lobe (ROI 18), showed more activation for No-go over Go in ten out of eleven subjects.

Stop-signal

We looked at group-level and individual means of δ across in condition in the SS task as well. In this paper, we will report only on the non-nuisance stop condition. However, we did also look at the nuisance distribution and compared it to the Go, No-go, and non-nuisance stop conditions. We found that its patterns and trends were very similar to the stop condition, and chose not to report it because it deviates from the canonical SS task.

From our group-level analysis, we found that many ROIs showed less activation after a Stop-signal than after a Go or No-go cue. Figure 9 shows aggregated group results for each ROI in the SS task. Each corresponds to a condition, with Go at the top, No-go in the middle, and Stop (again, not including the nuisance regressor) at the bottom. The methods for obtained this plot are identical from those used in the GNG analysis, but note that the scale in the percent BOLD legend are specific for this task, and ranges from -0.6 to 1.4, as opposed to the GNG task which ranges from -0.8 to 0.6. The upper bound is more than double, but our analyses will focus on the relative differences of activation in between conditions within a certain task. Go and No-go look very similar, with a few exceptions such as in the left insula (ROI 10). They especially look similar when compared to Stop, whose widespread cooler, negative activation looks starkly different from the wide ranges of Go/ No-go. The Stop condition has no ROIs with a percent change in BOLD signal larger than 0.21, and a mean of -0.036, which is much less than the means the ROIs for Go (mean = 0.40) and No-go (mean = 0.35). This is not to say that every ROI was less activated than every ROI in a Go or No-go condition, though, as we will find in the individual analysis.

The discussion of the individual analysis will focus on the ROIs where the means of the δ parameter were greater in one condition over another for all subjects. Like in GNG,

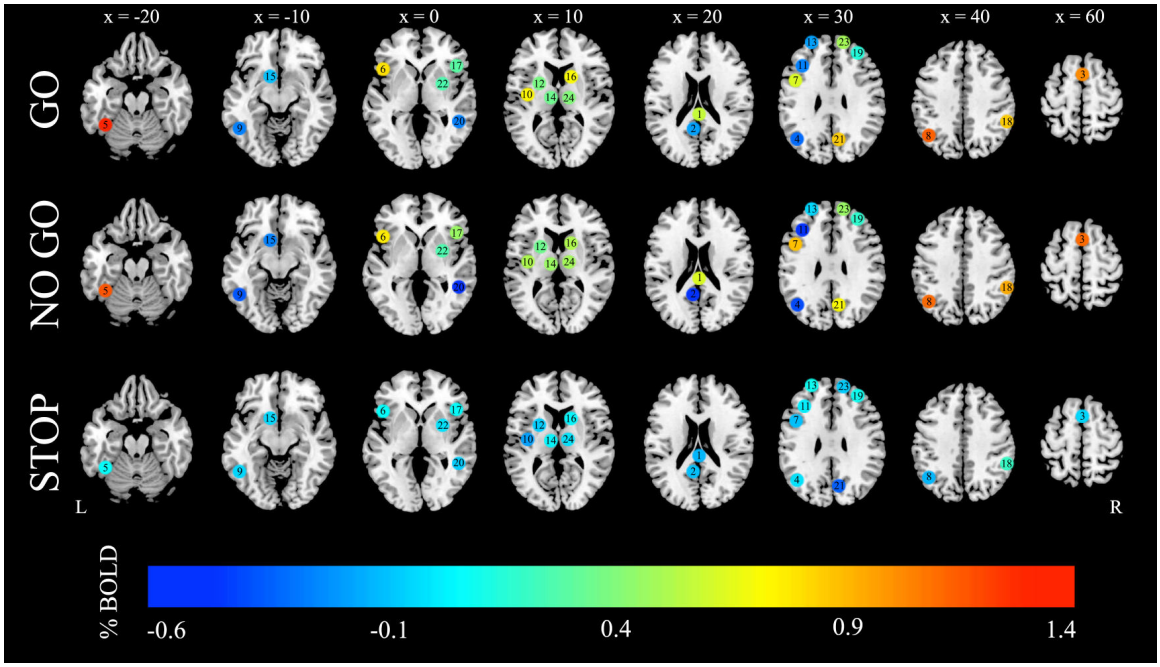

Figure 9. **Stop-Signal ROI Activation by Condition** Aggregated results from the SS task across individuals for the mean of the δ distribution (hyperparameter on the single-trial β) for each ROI. Cooler colors represent a smaller or more negative percent change in BOLD signal, i.e. less activation, and warmer colors represent a larger percent change in BOLD signal, i.e. more activation. The rows correspond to the mean of δ for each condition: Go, No-go, and Stop.

these analyses will focus on the results form the most complex model, Model 5. Unlike GNG, we have 3 different conditions to make comparisons about.

There were nine ROIs where all subjects showed more activation in Go than in Stop. They were: the preSMA (ROI 3), left fusiform gyrus (ROI 5), left IFG-1 (ROI 6), left IFG-2 (ROI 7), left IPL (ROI 8), left insula (ROI 10), right caudate (ROI 16), right IPL (ROI 18), and right SFG (ROI 23). In all of these ROIs except for the right IPL, this trend held true with $\pm$ two standard deviations from the mean of the posteriors. There were eight ROIs where all subjects showed more activation in No-Go than in Stop. These areas were the same as the ROIs with more activation in Go than in Stop listed previously, with the exception of the right caudate (which was more activated in No-Go than in Stop for ten out of eleven subjects). Only four of these eight areas (preSMA, left fusiform gyrus, left IFG-2, and left IPL) still showed more activation in No-Go than Stop with $\pm$ 2 standard

deviations from the mean of the posteriors.

Even though Figure 9 may give the impression that no ROIs showed more activation in the Stop condition than either Go or No-go, this is not the case. The left middle frontal gyrus (ROI 11) shows more activation for the Stop condition over the No-go condition in all eleven subjects. The left MFG was shown to be more active in stop than in Go for only seven out of eleven subjects, so the clear consensus was found only when comparing Stop and No-go. The left MFG was the only ROI with a consensus between all eleven subjects, but for ten out of eleven subjects, the left ITG (ROI 9) was found to be more activated in Stop than in *both* Go and No-go and the right MTG (ROI 20) was found to be more activated in Stop than in No-go.

There were no ROIs where all subjects showed more or less activation in the Go or No-go condition. However, there still is some evidence supporting the results from the GNG individual analysis. All of the areas found in the GNG task to have more activation in the Go condition than in the No-go condition (the left insula, left putamen, and left ventral striatum), were found again in the SS task to have more activation in the Go condition than in the No-go condition for at least nine out of the eleven subjects. However, the area found in the GNG task to have more activation in the No-go condition in the Go condition (the right inferior parietal lobe), was not found to follow this trend for a majority of subjects in the SS task.

Individual Differences in ROI Coactivation

As stated in the introduction, individual differences in cognitive regulation are thought to arise from differences in connectivity with the frontal cortex (Cole et al., 2012). Our more complicated models allow us to explore possible differences in connectivity. The parameter Σ in Models 4 and 5 show us which ROIs are coactivated. The twenty-four by twenty-four matrices shows pair-wise correlations of activation. Importantly, Σ does not specifically show connectivity, but coactivation. Thus, we cannot make any conclusions about connectivity, but coactivation and connectivity are closely related, so if the coactivation matrices

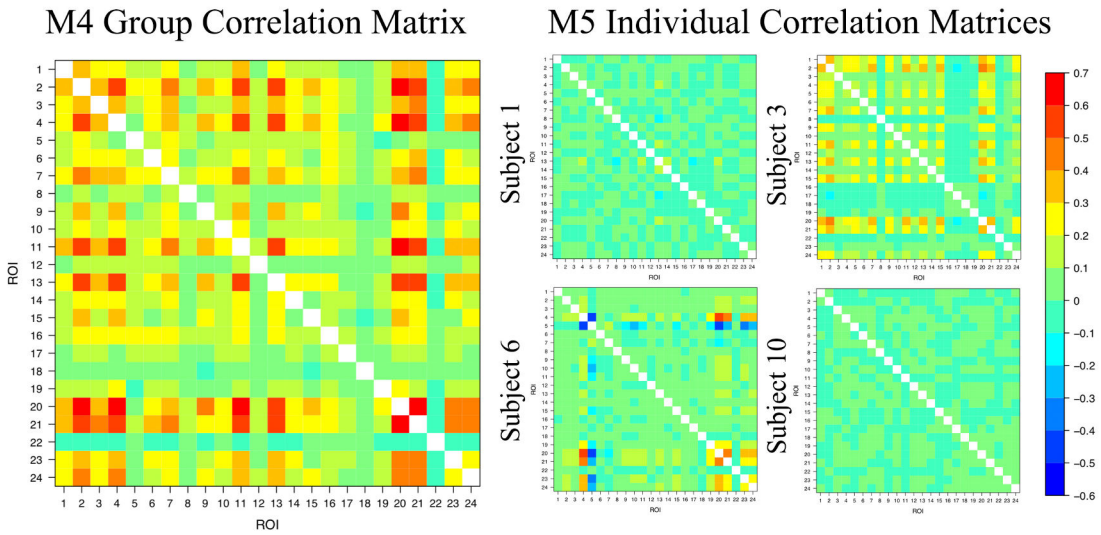

Figure 10. **GO/No-Go Correlation Matrices** Twenty-four by twenty-four correlation matrices showing coactivation of the twenty-four regions of interest. The correlation matrix on the left is the group correlation matrix from Model 4. The four correlation matrices on the right are four individual level-plots from Model 5. They are labeled with their respective subjects and represent the range of subjects. The legend to the right applies to all 5 plots. Cooler colors show a negative correlation and warmer colors show more positive correlations. The diagonal has a correlation of 1.0 and was removed for visual clarity.

for different subjects are vastly different, it is reasonable to suggest that the connectivity could also be different.

Go/ No-go

The model outputs Σ as a covariance matrix, but each prediction (for each sample and chain) was converted into a correlation matrix and then averaged and plotted. Figure 10 shows a plot of the Σ matrix estimated from Model 4 fit to GNG data and four plots of Σ_j estimated from Model 5 fit to GNG data from four representative subjects. In all five matrices, the diagonal components were removed to not skew the scale, since the diagonal represents the correlation between the same ROI (i.e. ROI 1 and ROI 1) and is always equal to 1.0. All five plots are colored according to the same scale. Warmer colors show a higher correlation and cooler colors show a smaller or more negative correlation.

This figure supports the idea that individual differences play a role in the GNG task.

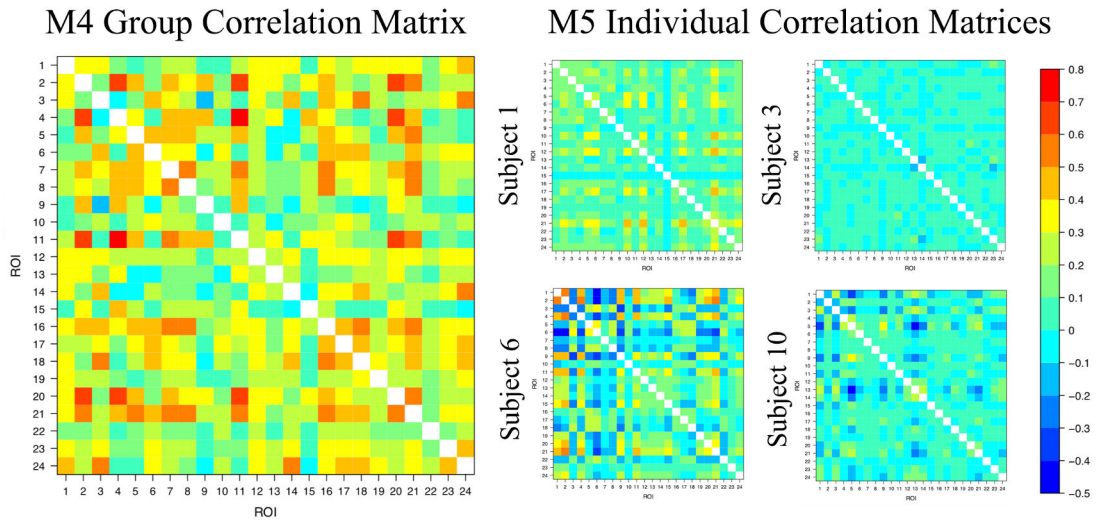

Figure 11. **Stop-Signal Correlation Matrices** Twenty-four by twenty-four correlation matrices showing coactivation of the twenty-four regions of interest. The correlation matrix on the left is the group correlation matrix from Model 4. The four correlation matrices on the right are four individual level-plots from Model 5.They are labeled with their respective subjects and represent the range of subjects. The legend to the right applies to all 5 plots. Cooler colors show a negative correlation and warmer colors show more positive correlations. The diagonal has a correlation of 1.0 and was removed for visual clarity.

Not only does the group level Σ not look like the individual plots, but the individual plots predicted by Model 5 also look very different. Subject 6 had the widest range of values in its mean Σ and greatly skewed the results. For the majority of subjects, the values of Σ ranged from approximately -0.2 to 0.2. However, even when fit to their own scale, no consistent pattern emerged.

Stop-signal

We also observed the Σ plots from Models 4 and 5 from the SS task, to see if these individual differences were again present and also to see if there are any similarities between the two tasks, on either a group or individual scale. Figure 11 shows the correlation matrix from Model 4 and representative figures from the correlation matrices from Model 5. It is important to note that the scale in this figure is for the ranges of the SS Σs and is not equivalent to the scale in Figure 10.

Figure 11 shows that evidence for individual differences were again observed in the SS task. The Σ estimated from Model 4 did not look like the Σ for each subject estimated from Model 5. Additionally, there was not a consistent pattern among the individual subjects' correlation matrices. Subject 6 again had the widest range of their correlation matrix, with the other subjects having much smaller ranges. Even when comparing across tasks, there was no consistent pattern emerging for either the group Σs in the SS or GNG tasks, or in individual Σs in the SS or GNG tasks.

Discussion

We used hierarchical Bayesian modeling to constrain and better understand fMRI data collected from GNG and SS tasks. First, we found evidence that hierarchical Bayesian modeling improves single-trial β estimates and that increasing levels of hierarchy improves estimates for the hyperparameters of single trail βs. However, we also found some contradictory results from Model 1. Model 1 very closely fit the observed neural time-series data, but had the least constrained single trial β estimates by far. This is contradictory, because the β estimates inform the neural predictions, so we would predict that if the neural predictions are close to the real data, the β estimates would be very constrained. It could be that Model 1 is overfitting the data, or even making the wrong predictions. Cross-validation is thus an important next step in determining the validity and generalizability of the model predictions.

Second, we found that the stopping condition had less BOLD activation across the brain than the Go or No-go conditions. This corresponds to many findings in the literature, but also contradicts some other evidence. The systematic deactivation in the stopping condition may be closely related to attention, specifically that the attention needed to initiate a stop response may cause the suppression of the default mode network (Turner, Van Maanen, & Forstmann, 2015). Additionally, Our results, however, do not provide evidence for the finding that the right inferior frontal cortex acts as a sort of brake in the brain (Aron, Robbins, & Poldrack, 2014). A possible reason for this disparity could be

because of the large number of voxels of the right IFC.

Third, we found strong evidence for individual differences in these tasks. This corresponds well to the literature, as individual differences are found to be at play in tasks of cognitive control, especially with response inhibition. The difference of the covariance matrices between tasks also suggest different neural systems are activated in the different tasks, which is supported by work in Swick et al. (2011).

While our analyses show some insight into neural dynamics, they do not reveal anything mechanistic or take the behavior into account. Accuracy of subjects and response times add more constraint to the model, and allow us to build in more theoretical components. To make mechanistic claims and to incorporate the behavior, a further direction of this research would be to build a full joint model (for a review see Turner, Forstmann, Love, Palmeri, and Van Maanen (2017)). However, even without the additional constraint of behavioral data, constructing a hierarchy allowed us to constrain the single-trial β estimates, discover diminished activation when stopping, and find evidence for individual differences in neural coactivation. In conclusion, we found that hierarchical Bayesian analysis is an important tool in understanding the neural dynamics of response inhibition.

Acknowledgements

Giwon Bahg, Xiangrui Li, and Zhong-Lin Lu were all collaborators on this project. This research was supported by Air Force Research Lab contract FA8650-16-1-6770.

References

Ahn, W.-Y., Krawitz, A., Kim, W., Busmeyer, J. R., & Brown, J. W. (2011). A model-based fmri analysis with hierarchical bayesian parameter estimation. *Journal of Neuroscience, Psychology, and Economics*, *4*, 95-110.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *18*, 177-185.

Bannon, S., Gonsalvez, C. J., Croft, R. J., & Boyce, P. M. (2002). Response inhibition deficits in obsessive-compulsive disorder. *Psychiatry Research*, *110*, 165-174.

Bowman, F. D., Caffo, B., Bassett, S. S., & Kilts, C. (2008). A bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage*, *39*, 146-156.

Cole, M. W., Yarkoni, T., Repovs, G., Anticevic, A., & Braver, T. S. (2012). Global connectivity of prefrontal cortex predicts cognitive control and intelligence. *The Journal of Neuroscience*, *32*, 8988-8999.

Jung, R., & Haier, R. (2007). The parieto-frontal integration theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30*, 135-154.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of the prefrontal cortex. *Annual Review of Neuroscience*, *24*, 167-202.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8-14.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving bold activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*, 2636-2643.

Palestro, J. J., Bahg, G., Sederberg, P. B., Lu, Z.-L., Steyvers, M., & Turner, B. M. (2018). *A tutorial on joint models of neural and behavioral measures of cognition.* (In Press)

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing.*

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage*, *23*, 752-763.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., ... Taylor, E. (2001). Mapping motor inhibition: Conjunctive brain activations across different versions of

go/no-go and stop tasks. *NeuroImage*, *13*, 250-261.

Schachar, R., & Logan, G. D. (1990). Impulsivity and inhibitory control in normal development and childhood psychopathology. *Developmental Psychology*, *23*, 710-720.

Swick, D., Ashley, V., & Turken, U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response inhibition tasks. *NeuroImage*, *56*, 1655-1665.

Turner, B. M., Forstmann, B. U., Love, B. U., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65-79.

Turner, B. M., Van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: The neural drift diffusion model. *Psychological Review*, *122*, 312-336.