Multiple brain activation patterns for the same task

Johan Nakuci[1*], Jiwon Yeon[2], Ji-Hyun Kim[3], Sung-Phil Kim[3] and Dobromir Rahnev[1]

[1]School of Psychology, Georgia Institute of Technology, Atlanta, Georgia, 30332, USA.
[2]Department of Psychology, Stanford University, Stanford, California, 94305, USA.
[3]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea.

*Corresponding author. Email: jnakuci3@gatech.edu

Keywords: perceptual decision making, clustering, fMRI, default mode network

Acknowledgments: This work was supported by the National Institutes of Health (grant R01MH119189 to DR) and the Office of Naval Research (grant N00014-20-1-2622 to DR).

Competing interests: Authors declare that they have no competing interests.

Author contributions:
Conceptualization: JN, DR
Methodology: JN, JY, DR
Data Curation: JY, JHK, SPK
Visualization: JN, DR
Funding acquisition: DR
Writing – original draft: JN, DR
Writing – review & editing: JN, JY, JHK, SPK, DR

Abstract

Meaningful variation in internal states that impacts cognition and behavior remains challenging to discover and characterize. Here we leveraged trial-to-trial fluctuations in the brain-wide signal recorded using functional MRI to test if distinct sets of brain regions are activated on different trials when accomplishing the same task. Subjects performed a perceptual decision-making task and provided confidence. We estimated the brain activations for each trial and clustered trials based on their similarity using modularity-maximization, a data-driven classification method. We identified three subtypes of trials that differed in both their activation profile and behavioral performance. Critically, Subtypes 1 and 2 were characterized by activations in different task-positive areas. Surprisingly, Subtype 3 exhibited strong activation in the default mode network, which is typically thought to decrease in activity during a task. Computational modeling revealed how the patterns of brain activity in each subtype emerged from interactions within and between large-scale brain networks. These findings demonstrate that the same task can be accomplished in the presence of widely varying brain activation patterns.

Introduction

Brain activity elicited by the same stimulus or task is highly variable[1,2]. Variation in task-induced brain activity has been identified in individual neurons[3] and large-scale brain networks[4] impacting cognition and behavior[5,6]. Variation in brain activity affects our actions in social situations[7], economic decisions[8], and even low-level perception[9].

Despite the widespread variability in brain activity during a task, standard analyses aim to identify the task-induced changes in brain activity across all trials[10]. Such analyses have been applied to a multitude of tasks such as face perception[11], memory[12], and navigation[13,14]. These standard analyses have identified a set of regions – termed "task-positive" – that increase in activity, and another set of regions – termed "task-negative" – that decrease in activity in response to external demands[15].

However, it is also possible that subsets of trials can produce meaningfully different patterns of activations that are not well captured by averaging across all trials. Indeed, the blood-oxygen level-dependent (BOLD) signal in fMRI is both spatially and temporally variable[16], with at least some of this variability likely to stem from meaningful variations in internal processing rather than simply noise[5,6,17,18]. Nevertheless, little attention has been devoted to investigating the different patterns of activations that could be elicited by the same task.

Here, we sought to identify discrete patterns of brain activity associated with the completion of the same perceptual decision-making task. To do so, we utilized a data-driven classification method to identify unique patterns in brain activity among individual trials. We found three different subtypes, with the first two showing activations in different task-positive sets of regions, while the third one surprisingly exhibiting task-negative activations. We further established the behavioral profile associated with each subtype, and used computational modeling to characterize how each subtype relates to activations within and the connectivity between established brain networks. Overall, our results indicate that multiple brain activation patterns can co-exist in the context of the same task.

Results

We examined the patterns of brain activation produced by a perceptual decision-making task where subjects judged the color (red or blue) with more dots and provided a confidence rating (**Fig. 1A**). We first performed standard GLM analyses to identify task-related brain activations. We observed increased activity in the visual and motor cortices, and decreased brain activity in the orbital frontal cortex and in the posterior cingulate cortex (**Fig. 1B**). However, single-trial beta responses estimated with a general linear model (GLM) using GLMsingle[47] deviated substantially from the group map. For example, unlike the group map, trial 2 for subject 1 showed strong positive activity in both the posterior cingulate cortex and the orbitofrontal cortex (**Fig. 1C**). On the other hand, trial 6 for subject 1 produced an activation pattern similar to the group map with negative activity in both the posterior cingulate cortex and the orbitofrontal cortex (**Fig. 1D**).

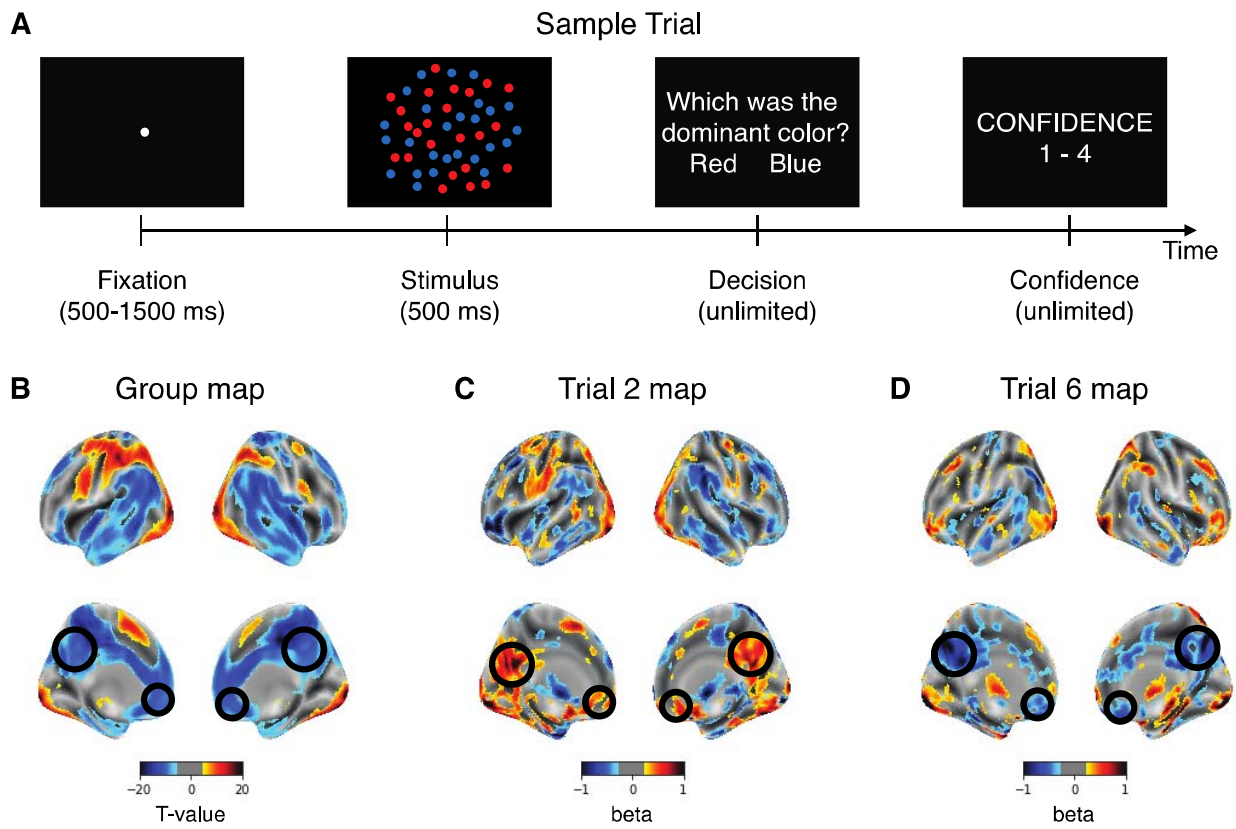

Figure 1. Variation in brain activity across individual trials. (A) Subjects performed a perceptual decision-making task that required them to judge the dominant color in a display of colored dots and rate their confidence. (B) Standard group analyses identified voxels with strong increases and decreases in task-induced brain activation across the cortex. The group brain map

is thresholded at $P < 0.001$ for display purposes. Brain activation maps for (C) trial 6 and (D) trial 2 from subject 1 demonstrate substantial variability across trials that is not represented in the standard group brain map. Brain maps for the individual trials are thresholded at $|beta| > 0.25$ for display purposes. Black circles highlight the activation in the posterior cingulate cortex and orbitofrontal cortex.

The divergence in brain activity between the two trials and to the group map highlights the possibility that more than one pattern of brain activation may exist when performing a task. To test for this possibility, we utilized a data-driven classification method to determine if multiple unique patterns of activation emerge across trials. For each trial we estimated the task-induced brain activity in each voxel and pooled trials across subjects ($N_{trials} = 35682$; $N_{subj} = 50$). We estimated the similarity across the activations between pairs of trials using Pearson correlation and clustered all trials using modularity-maximization to identify consistent activation patterns[19]. Clustering produced three subtypes of trials, each accounting for a roughly similar proportion of all trials. Specifically, Subtype 1 accounted for 34.25% of trials, Subtype 2 accounted for 35.61% of trials, and Subtype 3 accounted for 29.79% of trials (**Fig. 2A**). Further, each subtype was present in all subjects (**Fig. 2B**). To ensure the robustness of the subtypes, we performed a 2-fold cross-validation analysis using support vector machine (SVM) classifier. Specifically, we trained an SVM classifier on the subtype labels in half the data and tested its ability to predict the correct subtype labels in the remaining half of the data. We found that the SVM classifier correctly predicted subtype labels on 94.8% of the trials, thus confirming the robustness of the three subtypes (**Fig. 2C**).

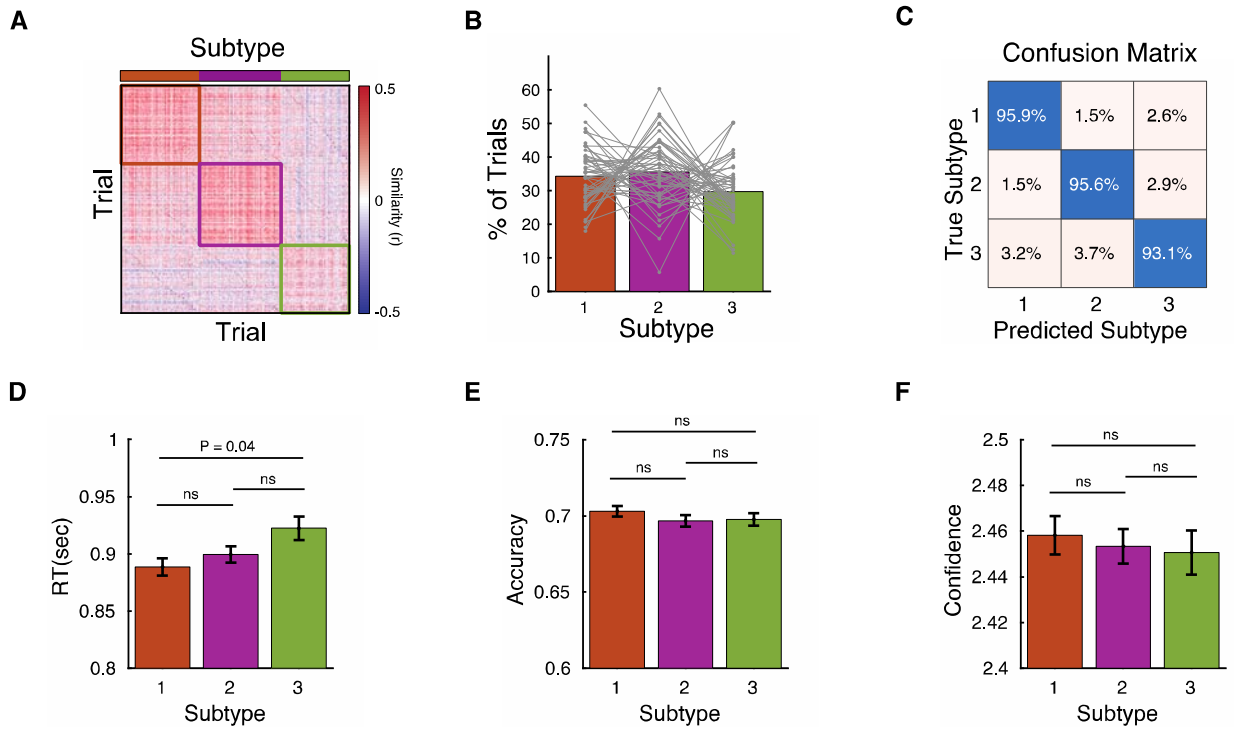

Figure 2. Subtypes of trials and behavioral performance. (A) Modularity-maximization clustering identified three subtypes of trials. The colored squares correspond to the trials composing each subtype. Pearson correlation was used to calculate the spatial similarity in activation (betas) among individual trials. Similarity matrix is ordered to show the stronger similarity of trials within subtypes, colored block. (B) The percent of trials classified as subtype 1, 2, or 3. The bar graphs represent the average proportion of trial across all individual. The dots represent the proportion of each trial subtype within an individual. (C) Confusion matrix showing SVM classification accuracy between true and predicted subtypes from a 2-fold cross-validation analysis. (D-F) Differences between subtypes in (D) reaction times (RT), (E) accuracy and (F) confidence. Statistical testing was conducted using paired-samples t-tests. ns, not significant.

Having identified these three different subtypes of trials, we compared how they differed in behavioral performance. We found that Subtype 3 featured longer reaction times compared to Subtype 1 ($t(49) = 2.10$, $P = 0.04$). However, Subtype 3 also featured numerically longer reaction times than Subtype 2 but the difference was not significant ($t(49) = 1.52$, $P = 0.14$; **Fig. 2D**). There was no significant difference between reaction times for Subtypes 1 and 2 ($t(49) = 0.90$, $P = 0.37$), or between any pair of subtypes in either accuracy ($P > 0.05$ for all three pairwise paired-samples t-tests; **Fig. 2E**) or confidence ($P > 0.05$ for all three pairwise paired-samples t-tests; **Fig. 2F**).

Critically, we examined the pattern of activation present in each subtype. To do so, we averaged the trials from each subtype within a subject. We then performed a group-level one-sample t-test on the average beta values to identify voxels with significant positive or negative activations (using a threshold of FDR-corrected $P < 0.01$; **Fig. 3**). We found that Subtype 1 was characterized by bilateral visual, parietal and left motor activations, as well as medial frontal, cingulate, temporal, and right motor deactivations. Subtype 2 had many similarities with Subtype 1 – such as visual and parietal activations couple with medial frontal, cingulate, and anterior temporal deactivations – but featured strong bilateral activations in the insula (whereas Subtype 1 had bilateral deactivations in the insula). Finally, Subtype 3 had the most surprising profile with strong activations along the default mode network and deactivations in a number of task-positive parietal and frontal areas. To further corroborate these results, we also performed a standard GLM with subtypes as factors in the regression and found similar patterns of activations (**Fig. S1**).

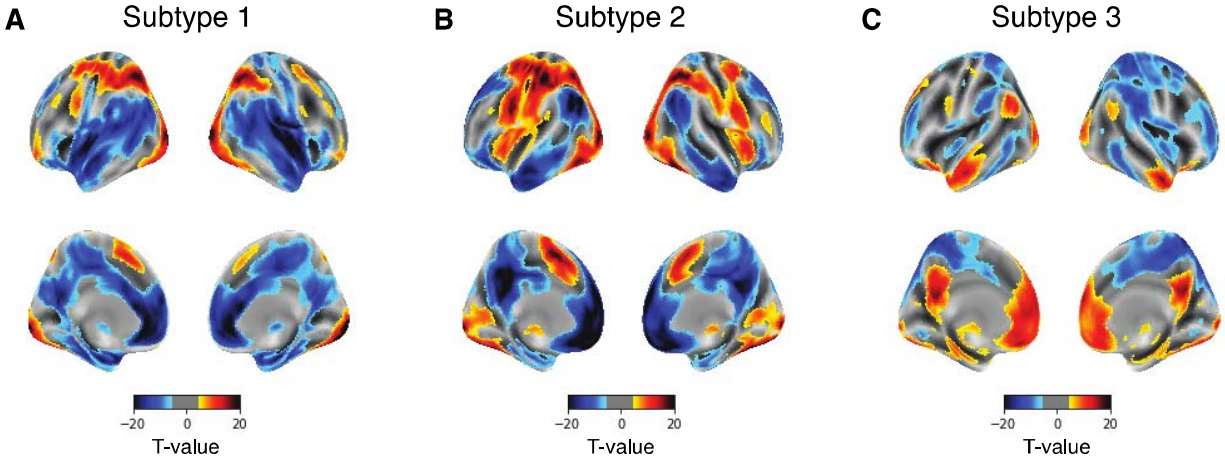

Figure 3. Brain activity associated with each subtype. Activation maps for (A) Subtype 1, (B) Subtype 2, and (C) Subtype 3, respectively. Each activation map was calculated by first averaging the trials for each subtype within a subject and followed by one-sample t-test to identify regions in which brain activity increased or decreased in response to the task. Brain maps are thresholded at $P_{FDR\text{-}corrected} < 0.01$.

To better characterize the differences between the three subtypes, we directly compared the patterns of activation among each pair of subtypes. We found that relative to Subtype 1, Subtype 2 exhibited strongly increased activity in bilateral insula, together with a host of additional

smaller differences in the frontal, parietal, and temporal lobes (paired-sample t-test; $P_{FDR\text{-}corrected} < 0.01$; **Fig. 4A**). On the other hand, compared to both Subtypes 1 and 2, Subtype 3 showed increased brain activity in the regions of the default mode network, including medial frontal cortex, posterior cingulate cortex, lateral parietal cortex, and anterior temporal cortex (paired-sample t-test; $P_{FDR\text{-}corrected} < 0.01$; **Fig. 4B, C**).

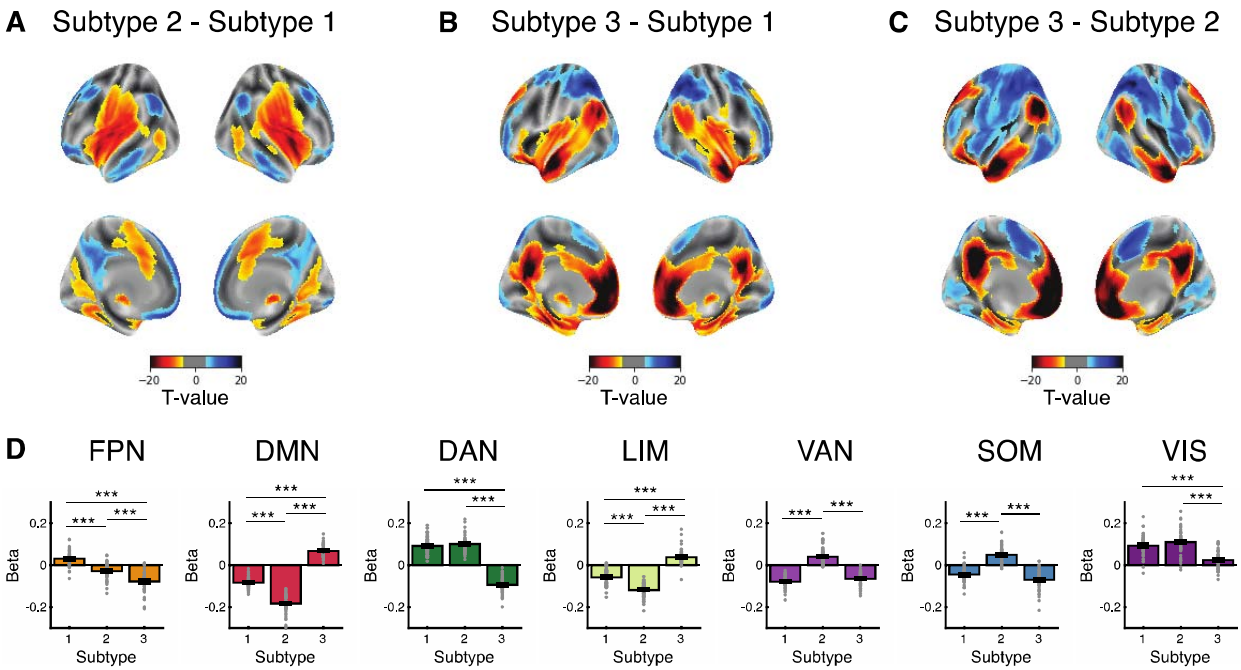

Figure 4. Activation differences between subtypes of trials. Brain activation differences between (A) Subtype 2 and Subtype 1, (B) Subtype 3 and Subtype 1, and (C) Subtype 3 and Subtype 2. (D) Activation for each of seven large-scale brain networks for each subtype. The bar graph shows the average change in activation (mean ± sem). The grey dots show the activation in each subject. Differences between subtypes were estimated using a paired-samples t-test. Brain maps are thresholded at $P_{FDR\text{-}corrected} < 0.01$. Network analysis, *** $P_{FDR\text{-}corrected} < 0.001$. FPN, Frontal Parietal Network; DMN, Default Mode Network; DAN, Dorsal Attention Network; LIM, Limbic; VAN, Ventral Attention Network; SOM, Somatomotor; VIS, Visual Network.

To further understand the nature of each subtype, we examined the pattern of activation within established brain networks for each subtype. For each subject we averaged the voxels within each of the seven brain networks – frontoparietal network (FPN), default mode network (DMN), dorsal attention network (DAN), limbic network, ventral attention network (VAN), somatomotor network, and visual network – defined in the Schaefer Atlas [20]. We found that Subtype 3

exhibited the strongest activations in DMN and the limbic networks but weakest activations in FPN, DAN, and the visual system (paired-samples t-tests; $P_{FDR\text{-}corrected}$; **Fig. 4D**). Compared to Subtype 2, Subtype 1 showed strong activation in FPN, DMN, and the limbic networks, but weaker activation in VAN and somatomotor networks. Interestingly, Subtypes 1 and 3 showed similar VAN and somatomotor activations, with both being significantly different from Subtype 2.

In addition to exploring the differences between the subtypes, we also examined what is common across them. To identify areas exhibiting consistent activations or deactivations in brain activity across the three subtypes, we identified the voxels in which the sign of activation was consistent among subtypes. Specifically, for each subtype, we binarized the brain maps for each subtype and summed them together to identify the areas of spatial overlap among all three subtypes. We found consistent activations in the visual and left motor cortex, as well as consistent deactivations in medial somatomotor, right motor, and bilateral temporal cortex (**Fig. 5**). Similar results were obtained when we examined the consistency of the maps at the subject level (**Fig. S2**).

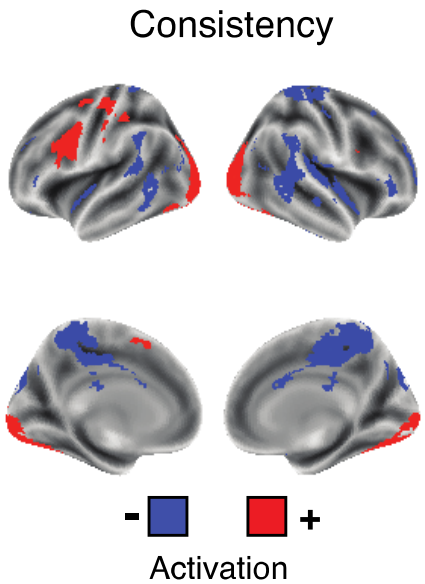

Figure 5. Consistency across the three subtypes. Voxels exhibiting consistent activation in all three subtypes. The activation maps for each subtype were binarized to identify the areas of spatial overlap among all three subtypes.

Finally, we sought to characterize the direct stimulus drive to each of the seven brain networks by fitting a computational model to the data. The critical idea behind the model is that observed activations in each network arise from two factors: (1) a direct stimulus drive to the network, (2) indirect contributions from other brain networks that are the product of their own stimulus drive times the strength of their structural connectivity with the given network (**Fig. 6A**). Specifically, the model assumes that observed activation, A_i, in network i can be expressed as:

$$A_i = \sum_{j=1}^{7} s_j \times w_{ij}$$

where s_j is the stimulus drive to network j, and w_{ij} is the structural connectivity strength between networks j and i. Note that the final activation for each network also depends on the within-network connectivity strength, w_{ii}. This approach is inspired by prior work on how interactions between large-scale brain networks influence the observed activity within each network[21–23]. We empirically estimated the structural connectivity from diffusion weighted imaging data of 50 subjects from the Human Connectome Project (see Methods). We then fit the model to the observed activations, A_i, to recover the hidden stimulus drives, s_i. Despite its simplicity, the model provided excellent fit to the activation data for each network (**Fig. 6B**).

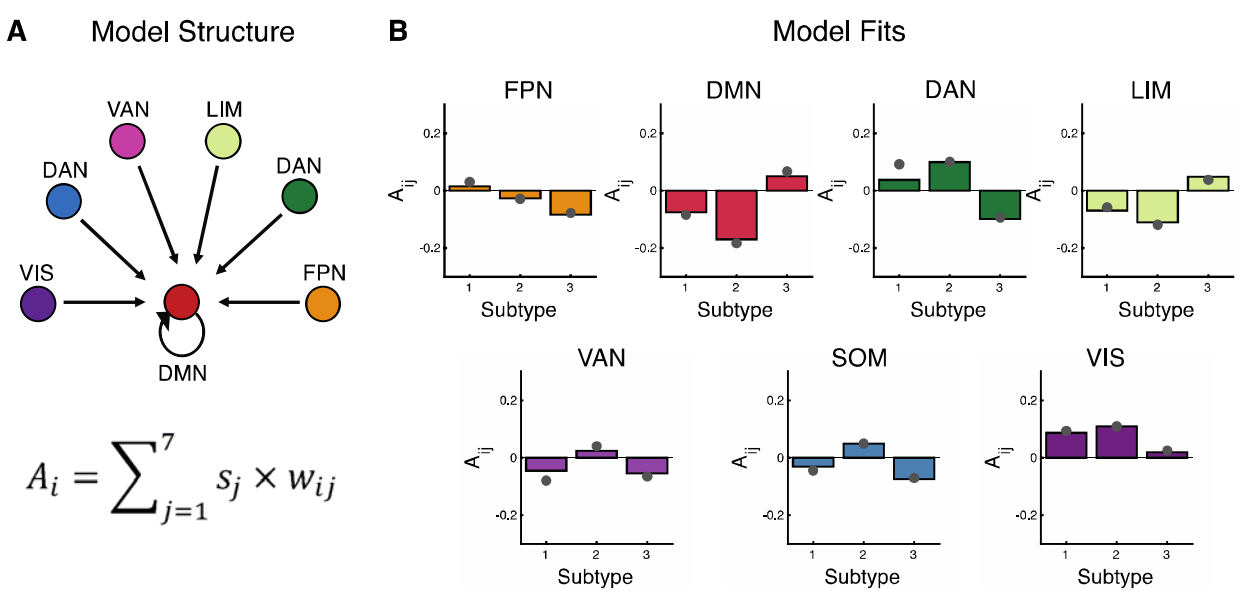

Figure 6. Model structure and fit. (A) Graphical depiction of the model. Activation (A_i) for a representative network, the DMN, is modeled as the product of the stimulus drive (s_j) and the strength structural connectivity between the DMN and all other networks (w_{ij}). (B) Group-level activation for each brain network across subtypes. The model was fit to the brain maps for each of the three subtype separately using the subject maps for fitting. Despite its simplicity, the model reproduces the empirical activation (gray dots) across brain networks very well.

Critically, the model allowed us to recover the strength of the stimulus drive to each network. The obtained stimulus drives mostly mirrored the observed stimulus activations, but with several important exceptions (**Fig. 7A-C**). First, the stimulus drive to the visual network was very similar across all subtypes even though we had observed stronger activations for Subtypes 1 and 2 compared to Subtype 3. Second, we found a very strong stimulus drive in VAN for Subtype 2, which suggests that one of the critical distinguishing factors of Subtype 2 may have been the presence of strong exogenous attention. Third, DAN showed consistently negative stimulus drive for all three subtypes even though it exhibited positive activations for Subtypes 1 and 2.

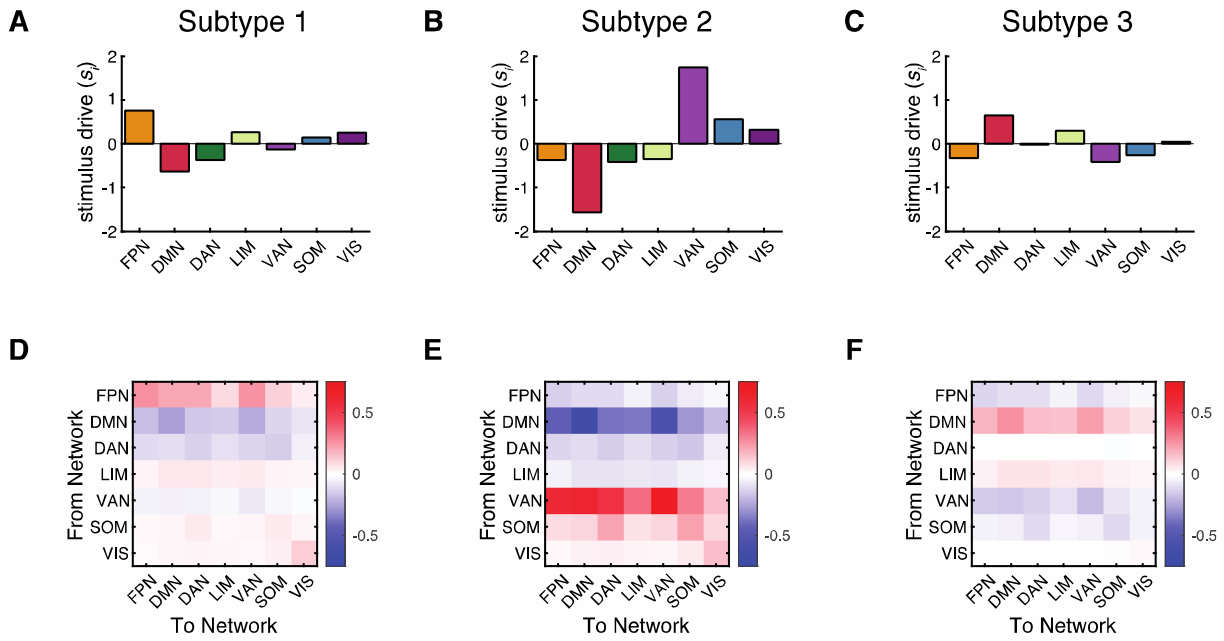

Figure 7. Stimulus drive and interactions among large-scale brain networks. The stimulus drive from model fitting in (A) Subtype 1, (B) Subtype 2, and (C) Subtype 3. The effect of the stimulus drive modulated by connectivity strength between networks for (D) Subtype 1, (E) Subtype 2, and (F) Subtype 3.

In addition, the model allowed us to gain an understanding of how interactions between brain networks effected the observed activation patterns (**Fig. 7D-F**). We found a dissociation between brain networks with the strongest stimulus-drive and activation. For Subtype 1 trials, reflecting the effect of the stimulus-drive, the FPN caused brain activity to increase globally, but the other networks tended to decrease the brain activity in the FPN (**Fig. 7D**). For Subtype 2 trials, the VAN increased brain activity in other networks despite exhibiting weak activation (**Fig. 7E**). Lastly, for Subtype 3 trials, the VAN and DMN exhibited the opposite pattern to Subtype 2 (**Fig. 7F**). In this case, the increased activation in the DMN induced other networks to also increase, whereas the VAN decreased brain activation globally. Interestingly, in Subtype 2, the DAN was strongly activated by VAN and somatomotor regions and deactivated by the DMN and limbic, and this interaction was reversed in Subtype 3. Taken together, these results suggest that a dynamic interaction exists between brain networks driven by the stimulus.

Discussion

Fluctuations in brain activity are ubiquitous during simple and cognitively demanding tasks[24,25] and are often indicative of variation in cognitive processing[26]. However, meaningful variation in internal states that impacts cognition and behavior remains challenging to discover and characterize. Here we examined if distinct patterns of brain activity emerge on different trials when accomplishing the same task. We utilized a data-driven clustering method based on modularity-maximation to identify consistent patterns of brain activity on individual trials. The clustering analysis identified three discrete subtypes of task-induced changes in brain activity. Surprisingly, while the first two subtypes mapped onto complementary task-positive networks, the third subtype aligned best with the task-negative network. Further, computational modeling suggested that the activation differences between the subtypes arose from alterations in the stimulus representation across large-scale brain networks. These findings demonstrate that the same task can be accomplished in the presence of widely varying brain activation patterns.

Our most striking finding was that trials from Subtype 3 showed a strong increase in DMN activity despite the fact that subjects were engaged in a perceptual decision-making task that canonically leads to DMN suppression. Behaviorally, Subtype 3 was characterized by longer

RTs, but the difference with Subtype 1 (34 ms; P = 0.04) and Subtype 2 (23 ms; P = 0.14) were very small. There were also no differences in accuracy or confidence. These behavioral results are particularly surprising because DMN activation is usually associated with mind-wandering and being off-task[27,28], yet such interpretations would presumably predict larger behavioral differences between Subtype 3 and Subtypes 1/2. Our findings demonstrate substantial DMN activations even for tasks that canonically should deactivate DMN, which challenges the notion that DMN should be viewed as purely task-negative network[29].

Both Subtype 1 and Subtype 2 mapped well to traditional task-positive brain areas. However, a critical difference between the two was the presence of strong insula activation in Subtype 2 and a strong insula deactivation in Subtype 1. At the network level, there were qualitative differences where Subtype 2 showed negative activations in FPN and positive activations in VAN and somatomotor regions, whereas Subtype 1 showed the opposite pattern. FPN is thought to play a significant role in the top-down modulation[30–32] and conflict monitoring[33,34], whereas VAN is involved in detecting the salient component of the stimulus to guide attention and behavior[35,36]. These differences in activations are consistent with the notion that Subtype 2 is linked more strongly to exogenous attention, whereas Subtype 1 is linked to endogenous attention.

Why are three qualitatively different patterns of brain activity emerging from different trials of the same task? One may speculate that a cognitive process can be accomplished through multiple pathways. Similar to a person driving home from work, there may be multiple routes to get home and the path taken may depend on various factors such as the extent of traffic on a specific route. In a similar manner, multiple brain pathways may allow for the completion of a given task. In this interpretation, Subtypes 1 may be characterized as a decision-making pathway in the presence of endogenous attention, Subtype 2 as a decision-making pathway in the presence of exogenous attention, and Subtype 3 as a decision-making pathway for internally focused attention. Of course, instead of discrete pathways, another possibility is that the trial subtypes are part of a gradient in activation across the brain[37,38]. In this interpretation, the trial subtypes would indicate the axes of this space, and the location of an individual trial within this space could be dependent on the extent of top-down control[39,40], the salience of the stimulus[41], and internally oriented attention[42,43]. Critically, neither interpretation excludes the existence of shared

components in the decision-making processes. Indeed, we found that the visual and motor regions exhibited similar patterns in activation in all subtypes. Additionally, this consistent pattern of activation included regions in the frontal, parietal, and temporal areas, suggesting a core component across all trials. Future research should examine the underlying processes that lead to the emergence of these three trial subtypes.

Our computational model allowed us to break down the observed activations in each of the seven major brain networks as a combination of the stimulus drive directly to each network and the interactions between the networks. The model confirmed that, as would be expected, the stimulus drive to the visual network was similar across all three subtypes despite differences in observed activations, which the model explained as arising from interactions with other networks. Another noteworthy result of the modeling was the strong stimulus drive in VAN, a network associated with exogenous attention and saliency detection[35,44,45], for Subtype 2 relative to the observed activation. In addition, the model can be used to track the complex interactions across networks. These results demonstrate that our computational model was able to uncover mechanisms that would not have been apparent in its absence.

In conclusion, we found that the same perceptual decision-making task featured three distinct patterns of brain activation with only minimal behavioral differences between them. These results suggest that multiple independent pathways may be employed to accomplish a task.

Methods

Dataset

Our analysis utilized a dataset consisting of 50 healthy subjects (25 females; Mean age = 26; Age range = 19-40) that has been described elsewhere[46]. All subjects were screened for any history of neurological disorders or MRI contraindications. The study was approved by Ulsan National Institute of Science and Technology Review Board (UNISTIRB-20-30-C) and all subjects gave written consent.

Task

Subjects had to determine which set of colored dots (red vs. blue) is more frequent in a cloud of dots. Each trial began with a white fixation dot presented for a variable amount of time between 500-1500 ms. Then, the stimulus was shown for 500 ms, followed by untimed decision and confidence screens. The stimulus consisted of between 140 and 190 red- and blue-colored dots (dot size = 5 pixels) dispersed randomly inside an imaginary circle with a radius of 3° from the center of the screen. Four different dot ratios were used – 80/60, 80/70, 100/80, and 100/90, where the two numbers indicate the number of dots from each color. Subjects performed a total of 768 trials. Three subjects completed only half of the 6th run and another three subjects completed only the first 5 runs due to time constraints. The remaining 44 subjects completed the full 6 runs.

MRI Recording

The MRI data was collected on a 64-channel head coil 3T MRI system (Magnetom Prisma; Siemens). Whole-brain functional data were acquired using a T2*-weighted multi-band accelerated imaging (FoV = 200 mm; TR = 2000 ms; TE = 35 ms; multiband acceleration factor = 3; in-plane acceleration factor = 2; 72 interleaved slices; flip angle = 90°; voxel size = 2.0 x 2.0 x 2.0 mm^3). High-resolution anatomical MP-RAGE data were acquired using T1-weighted imaging (FoV = 256 mm; TR = 2300 ms; TE = 2.28 ms; 192 slices; flip angle = 8°; voxel size = 1.0 x 1.0 x 1.0 mm^3).

MRI Preprocessing

MRI data were preprocessed with SPM12 (Wellcome Department of Imaging Neuroscience, London, UK). We first converted the images from DICOM to NIFTI and removed the first three volumes to allow for scanner equilibration. We then preprocessed with the following steps: de-spiking, slice-timing correction, realignment, segmentation, coregistration, normalization, and spatial smoothing with 10 mm full width half maximum (FWHM) Gaussian kernel. Despiking was done using the 3dDespike function in AFNI. The preprocessing of the T1-weighted structural images involved skull-removal, normalization into MNI anatomical standard space, and segmentation into gray matter, white matter, and cerebral spinal fluid, soft tissues, and air and background.

Single-Trial Beta Estimation

Single-trial beta responses were estimated with a general linear model (GLM) using GLMsingle, a Matlab toolbox for single-trial analyses[47]. The hemodynamic response function was estimated for each voxel and nuisance regressors were derived in the same manner as previously described in Allen et al.[48]. Additionally, regressors for the global signal and for six motion parameters (three translation and three rotation) were included. The single-trial betas were estimated in three batches. In each batch, the betas for every third trial were estimated because the trials in our study were temporally close together. Also, trials that were within 20 seconds from the end of run were removed. The betas for each voxel represent the estimated trial-wise BOLD response and are relative to the BOLD signal observed during the absence of the stimulus[47].

Modularity-maximization Based Clustering

The beta maps per trials were pooled among subjects to ensure that there was consistency in clustering correspondence. A trial-by-trial similarity matrix was created using the Pearson Product Correlation. Clustering of the similarity matrix was conducted using modularity-maximization[19]. Modularity-maximization does not require the number of clusters to be specified and the resolution of the clusters was controlled with resolution parameter, $\gamma = 1$. Modularity-maximization was implemented with the Generalized Louvain algorithm part of the Brain Connectivity Toolbox[49]. Clustering results from the Generalized Louvain algorithm can be dependent on the initial seeding. Therefore, clustering was repeated 100x with different initial seeds, and final clustering results were based on the consensus across iterations[50]. Specifically,

consensus clustering identifies a single representative partition from the set of 100 iterations. This process involves the creation of a thresholded nodal association matrix which describes how often nodes were placed in the same cluster. The representative partition is then obtained by using a Generalized Louvain algorithm to identify clusters within the thresholded nodal association matrix[50]. We have previously utilized this method to identify trial subtypes in brain activity measured with electroencephalography in a motion discrimination[51] task and working memory task[52].

Standard Group-Level Analysis

A standard task-based GLM analysis was conducted to identify voxels in which the beta values significantly deviated from zero. Specifically, a single activation brain map was created per subject by averaging the individual beta maps across trials and a one-sample t-tests was conducted across subjects to identify the regions that deviated from zero.

Trial Subtype Activation

In a similar manner to the standard group-level analysis, a trial subtype task-based analysis was conducted to identify voxels in which the beta values for each subtype of trials significantly deviated from zero. For each subtype, a single activation brain map was created per subject by averaging the individual beta maps across trials in each subtype. A group-level one-sample t-tests was conducted to identify voxels with significant increase or decrease in activation. Activation maps were false discovery rate (FDR) corrected for multiple comparison.

Standard General Linear Modeling

To further corroborate the results, we also performed a standard GLM using SPM12 with subtypes as factors in the regression and found similar patterns of activations. We fit a GLM that allowed us to estimate the beta values for each voxel in the brain in each of the trial subtypes. The model consisted of separate regressors for each of the subtypes, inter-block rest periods, as well as linear and squared regressors for six motion parameters (three translation and three rotation), five tissue-related regressors (gray matter, white matter, and cerebrospinal fluid), global signal, and a constant term per run.

Consistency in Activation Across Subtypes

To identify voxels exhibiting consistent task-induced changes in brain activity, we examined the consistency of the sign of voxel activations (positive or negative) across subjects. To do so, for each of the subtype brain maps, we first removed all non-gray matter voxels. We then binarized the voxel activation values $activation_i$ such that:

$$binary_i = \begin{cases} 1, & activation_i > 0 \\ -1, & activation_i < 0 \end{cases}$$

The consistency of the sign of a voxel's activation across subtypes (C_i) was then calculated as total number of subtypes for a which voxel i was positively or negatively activated using the formula:

$$C_i = \sum_{i=1}^{3} binary_i$$

As defined, C_i goes from -3 (all subtypes having negative activation for that voxel) to 3 (all subtypes having positive activations for that voxel). C_i was then thresholded at ± 3 to identify voxels that consistently increased or decreased in brain activity across subtypes. The same analysis was repeated within each subject.

Voxel and Brain Network Differences Between Subtypes

Differences in task-based brain activity between subtypes was conducted to identified voxels in which the beta values between subtypes differed. The analysis was conducted both at the voxelwise and between large-scale brain networks. For voxelwise analyses, a paired t-test was used to test for differences between subtypes. For the comparison between large-scale brain networks, the beta values from voxels associated with one of seven large-scale brain networks part of the Schaefer Atlas[20] were averaged together within a subject and a paired t-test was used to test for differences between subtypes. All p-values were false discovery rate (FDR) corrected for multiple comparison.

Behavioral Performance Differences Between Subtypes

Paired-samples t-test was used to test for differences in accuracy, reaction times and confidence between subtypes. Further, we tested for behavioral differences within each condition and for the composition subtypes within each condition (**Fig. S3**).

SVM Classification

A 2-fold cross-validation analysis using Support Vector Machine (SVM) classifier was performed with MATLAB's *fitcecoc.m.* Trials were randomly separated into two bins containing 50% of trials each. The SVM classifier was trained on the first bin of trials and tested on the second bin. The procedure was repeated with the training and testing bins swapped. Classification accuracy was computed by comparing the predicated to the empirical trial subtype label.

Model Specification

We sought to characterize the hidden stimulus drive to each of the seven brain networks by fitting a computational model to the data. The model generates brain activity among large-scale brain networks using minimal assumptions. This approach is inspired by prior work on examining the influence interactions between large-scale brain networks have on the activity of an individual brain network[21–23].

The critical idea behind the model is that observed activations in each network arise from two factors: (1) a direct stimulus drive to the network, (2) indirect contributions from other brain networks that are the product of their own stimulus drive times the strength of their structural connectivity with the given network. Specifically, the model assumes that observed activation, A_i, in network i can be expressed as:

$$A_i = \sum_{j=1}^{7} s_j \times w_{ij}$$

where s_j is the stimulus drive to network j, and w_{ij} is the structural connectivity strength between networks i and j. Note that the final activation for each network also depends on the within-network connectivity strength, w_{ii}. For visualization purposes s_j is divided by 1000.

Structural Connectivity

Structural connectivity between brain regions was estimated from 50 healthy human data from the S1200 release from the Human Connectome Project (HCP) [53]. Fiber tracking was conducted

using DSI Studio with a modified FACT algorithm[54]. Diffusion MRI images were reconstructed in MNI space using generalized q-sampling imaging and fiber tracking was performed until 250,000 streamlines were reconstructed with angular threshold of 50°, step size of 2 mm, minimum length of 10mm, and maximum length of 400mm.

For each subject, an undirected weighted structural connectivity matrix, w, containing 200 brain regions part of the Schaefer atlas was constructed by counting the number of streamlines between brain regions[20]. Connectivity matrices were then normalized by dividing the number of streamlines (T) by the combined volumes (v) of regions i and j:

$$W_{ij} = \frac{T_{ij}}{v_i + v_j}$$

The strength of the connectivity between large-scale brain networks was estimated by first averaging the matrices, w, across subjects to generate a group structural connectivity map followed by averaging the strength of the connectivity between regions associated between the seven large-scale brain networks in the Schaefer Atlas. This resulted in a 7x7 matrix where each entry represents the strength of the connectivity within and between brain networks.

Model Fitting

We first fit the model to the empirically observed beta values from the 50 subjects separately for each of the subtypes. The model had seven free parameters corresponding to the strength of the stimulus drive, s_i, in each of the brain networks. We simulated the activation for each of the seven networks and fitting minimized the mean squared error (MSE) between the simulated and empirically observed values. Once the fitting was completed, we also generated the predictions of the best-fitting model for the activation values for each of the large-scale brain networks per subtype. The fitting itself was carried out using the Bayesian Adaptive Direct Search (BADS) toolbox[55]. We fit the model 25 times and we reported the best fitting model among the 25 iterations.

Data and Code Availability

All data and code will be made available upon publication.

References

1. Arieli, A., Sterkin, A., Grinvald, A. & Aertsen, A. Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science* (1996) doi:10.1126/science.273.5283.1868.
2. Goris, R. L. T., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nat Neurosci* (2014) doi:10.1038/nn.3711.
3. Ribrault, C., Sekimoto, K. & Triller, A. From the stochasticity of molecular processes to the variability of synaptic transmission. *Nature Reviews Neuroscience* Preprint at https://doi.org/10.1038/nrn3025 (2011).
4. Clare Kelly, A. M., Uddin, L. Q., Biswal, B. B., Castellanos, F. X. & Milham, M. P. Competition between functional brain networks mediates behavioral variability. *Neuroimage* (2008) doi:10.1016/j.neuroimage.2007.08.008.
5. Reinhart, R. M. G. & Nguyen, J. A. Working memory revived in older adults by synchronizing rhythmic brain circuits. *Nat Neurosci* **22**, 820–827 (2019).
6. Mišić, B., Mills, T., Taylor, M. J. & McIntosh, A. R. Brain Noise Is Task Dependent and Region Specific. *J Neurophysiol* **104**, 2667–2676 (2010).
7. Durlauf, S. N. A framework for the study of individual behavior and social interactions. *Sociol Methodol* **31**, (2001).
8. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nat Neurosci* **10**, (2007).
9. Afraz, A., Pashkam, M. V. & Cavanagh, P. Spatial heterogeneity in the perception of face and form attributes. *Current Biology* **20**, (2010).
10. Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C. & Worsley, K. J. Multisubject fMRI Studies and Conjunction Analyses. *Neuroimage* **10**, 385–396 (1999).
11. Kanwisher, N., McDermott, J. & Chun, M. M. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience* **17**, 4302 LP – 4311 (1997).
12. Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R. & Haynes, J.-D. The Distributed Nature of Working Memory. *Trends Cogn Sci* **21**, 111–124 (2017).
13. Patai, E. Z. & Spiers, H. J. The Versatile Wayfinder: Prefrontal Contributions to Spatial Navigation. *Trends Cogn Sci* **25**, 520–533 (2021).
14. Spiers, H. J. & Maguire, E. A. A navigational guidance system in the human brain. *Hippocampus* **17**, 618–626 (2007).
15. Fox, M. D. *et al.* The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A* (2005) doi:10.1073/pnas.0504136102.
16. Garrett, D. D. *et al.* Moment-to-moment brain signal variability: A next frontier in human brain mapping? *Neurosci Biobehav Rev* **37**, 610–624 (2013).
17. Waschke, L., Wöstmann, M. & Obleser, J. States and traits of neural irregularity in the age-varying human brain. *Sci Rep* **7**, 17381 (2017).
18. McIntosh, A. R., Kovacevic, N. & Itier, R. J. Increased Brain Signal Variability Accompanies Lower Behavioral Variability in Development. *PLoS Comput Biol* **4**, e1000106- (2008).

19. Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. & Onnela, J. P. Community structure in time-dependent, multiscale, and multiplex networks. *Science* (2010) doi:10.1126/science.1184819.
20. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex* **28**, 3095–3114 (2018).
21. Cole, M. W., Ito, T., Bassett, D. S. & Schultz, D. H. Activity flow over resting-state networks shapes cognitive task activations. *Nat Neurosci* **19**, 1718–1726 (2016).
22. Gu, S. *et al.* Controllability of structural brain networks. *Nat Commun* **6**, 8414 (2015).
23. Honey, C. J. & Sporns, O. Dynamical consequences of lesions in cortical networks. *Hum Brain Mapp* (2008) doi:10.1002/hbm.20579.
24. Garrett, D. D. *et al.* Moment-to-moment brain signal variability: A next frontier in human brain mapping? *Neurosci Biobehav Rev* **37**, 610–624 (2013).
25. Faisal, A. A., Selen, L. P. J. & Wolpert, D. M. Noise in the nervous system. *Nat Rev Neurosci* **9**, 292–303 (2008).
26. Waschke, L., Kloosterman, N. A., Obleser, J. & Garrett, D. D. Behavior needs neural variability. *Neuron* **109**, 751–766 (2021).
27. Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N. & Andrews-Hanna, J. R. Mind-wandering as spontaneous thought: a dynamic framework. *Nat Rev Neurosci* **17**, 718–731 (2016).
28. Mason, M. F. *et al.* Wandering Minds: The Default Network and Stimulus-Independent Thought. *Science (1979)* **315**, 393–395 (2007).
29. Spreng, R. N. The Fallacy of a “Task-Negative” Network. *Frontiers in Psychology* vol. 3 Preprint at https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00145 (2012).
30. Zanto, T. P., Rubens, M. T., Thangavel, A. & Gazzaley, A. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat Neurosci* **14**, 656–661 (2011).
31. Liu, Y., Bengson, J., Huang, H., Mangun, G. R. & Ding, M. Top-down Modulation of Neural Activity in Anticipatory Visual Attention: Control Mechanisms Revealed by Simultaneous EEG-fMRI. *Cerebral Cortex* **26**, 517–529 (2016).
32. Gazzaley, A. & D’Esposito, M. Top-Down Modulation and Normal Aging. *Ann N Y Acad Sci* **1097**, 67–83 (2007).
33. Dosenbach, N. U. F. *et al.* Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences* **104**, 11073–11078 (2007).
34. Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W. & Schacter, D. L. Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *Neuroimage* **53**, 303–317 (2010).
35. Sridharan, D., Levitin, D. J. & Menon, V. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc Natl Acad Sci U S A* **105**, 12569–12574 (2008).
36. Seeley, W. W. *et al.* Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience* **27**, 2349–2356 (2007).

37. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences* **113**, 12574–12579 (2016).
38. Tian, Y., Margulies, D. S., Breakspear, M. & Zalesky, A. Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nat Neurosci* **23**, 1421–1432 (2020).
39. Cole, M. W. *et al.* Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci* **16**, 1348–1355 (2013).
40. Koechlin, E., Ody, C. & Kouneiher, F. The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science (1979)* **302**, 1181–1185 (2003).
41. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* **3**, 201–215 (2002).
42. Rosenberg, M. D. *et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nat Neurosci* **19**, (2015).
43. Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M. & Valera, E. M. Dynamic Brain Network Correlates of Spontaneous Fluctuations in Attention. *Cerebral Cortex* **27**, 1831–1840 (2017).
44. Menon, V. & Uddin, L. Q. Saliency, switching, attention and control: a network model of insula function. *Brain structure & function* vol. 214 655–667 Preprint at https://doi.org/10.1007/s00429-010-0262-0 (2010).
45. Seeley, W. W. The salience network: A neural system for perceiving and responding to homeostatic demands. *Journal of Neuroscience* **39**, 9878–9882 (2019).
46. Nakuci, J. *et al.* Quantifying the idiosyncrasy in brain activations. *bioRxiv* 2022.08.01.502338 (2023) doi:10.1101/2022.08.01.502338.
47. Prince, J. S. *et al.* Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *Elife* **11**, e77599 (2022).
48. Allen, E. J. *et al.* A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat Neurosci* **25**, 116–126 (2022).
49. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
50. Bassett, D. S. *et al.* Robust detection of dynamic community structure in networks. *Chaos* (2013) doi:10.1063/1.4790830.
51. Nakuci, J., Samaha, J. & Rahnev, D. Brain signatures indexing variation in internal processing during perceptual decision-making. *bioRxiv* 2023.01.10.523502 (2023) doi:10.1101/2023.01.10.523502.
52. Nakuci, J., Covey, T. J., Shucard, J. L., Shucard, D. W. & Muldoon, S. F. Single trial variability in neural activity during a working memory task reveals multiple distinct information processing sequences. *Neuroimage* **269**, 119895 (2023).
53. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *Neuroimage* (2013) doi:10.1016/j.neuroimage.2013.05.041.
54. Yeh, F. C., Verstynen, T. D., Wang, Y., Fernández-Miranda, J. C. & Tseng, W. Y. I. Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PLoS One* **8**, 1–16 (2013).

55. Acerbi, L. & Ma, W. J. Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search. *In Advances in Neural Information Processing Systems* **30**, 1834–1844 (2017).

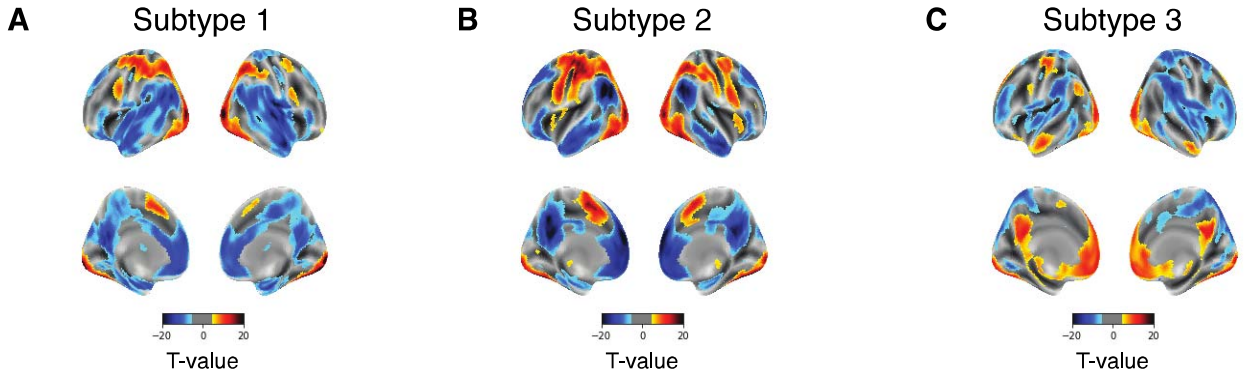

Figure S2. Standard GLM analysis with trial subtype labels as factors in the regression. Traditional GLM analysis was able to reproduce the changes in brain activity for (A) Subtype 1, (B) Subtype 2, and (C) Subtype 3. Brain maps are thresholded at FDR < 0.01.

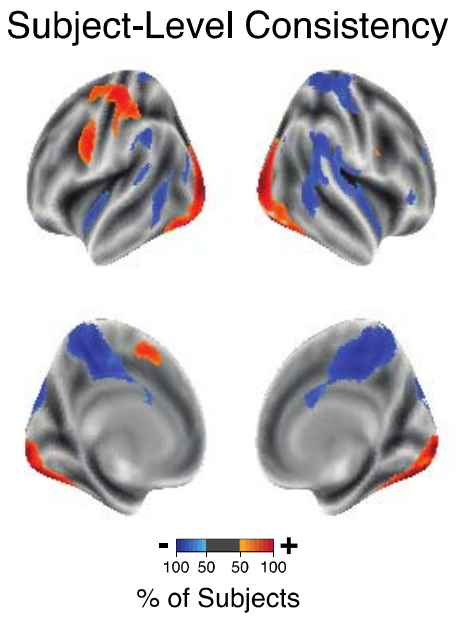

Figure S2. Subject-level consistency across the three subtypes. For each subject, voxels exhibiting consistent activation in all three subtypes were identified by binarizing the activation maps for each subtype and summing together to identify the areas of spatial overlap among all three subtypes. Consistency maps were then averaged across subjects and a voxel had to be consistent in 50% of subjects to be considered.

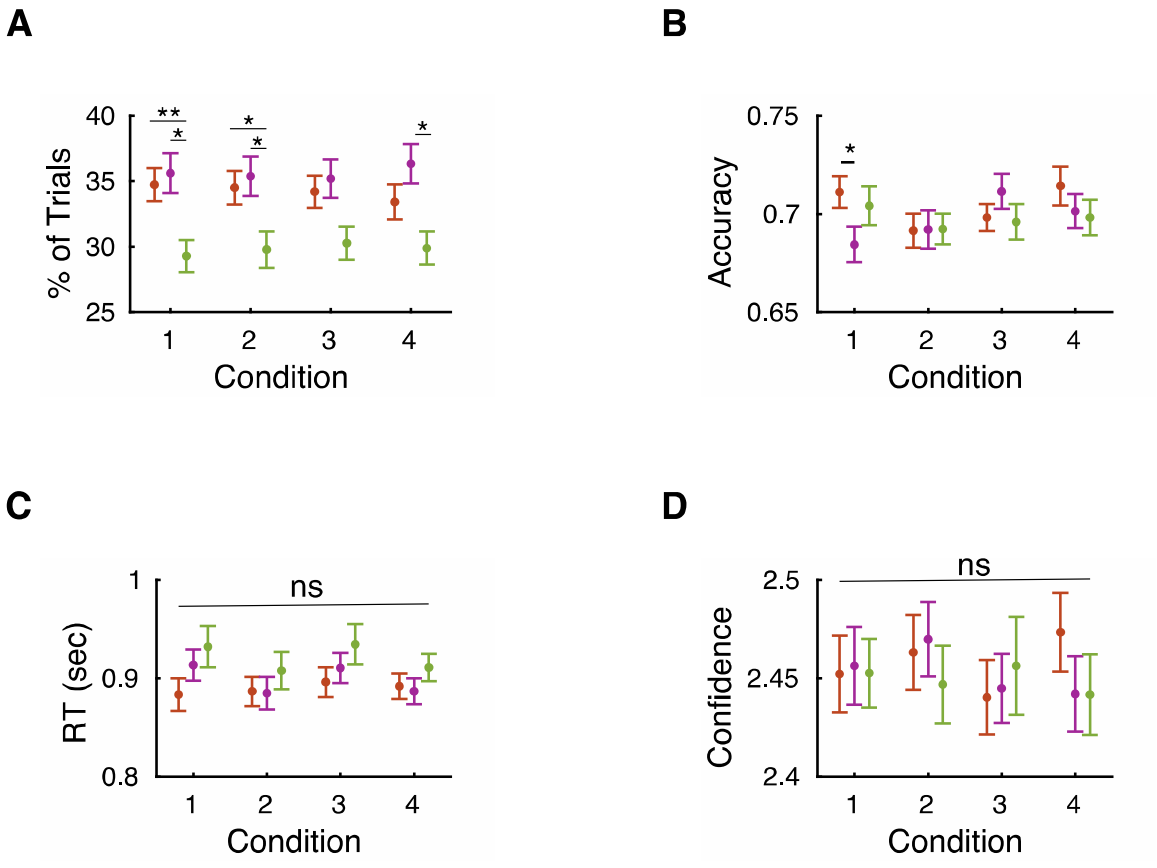

Figure S3. Subtype differences in behavior performance within each condition. (A) Differences between subtypes in (A) percentage of trials for each condition. Critically all subtypes are present in every condition indicating that the clustering is not simply identifying differences the different conditions. Differences between subtypes in (B) accuracy; (C) reaction times (RT); and (D) confidence for each condition. Statistical testing was conducted using paired-samples t-test. **P < 0.01; *P < 0.05; ns = not significant.