PSYCHOMETRIKA—VOL. 88, NO. 2, 434–455
JUNE 2023
https://doi.org/10.1007/s11336-023-09907-8

BLIND SUBGROUPING OF TASK-BASED FMRI

ZACHARY F. FISHER

THE PENNSYLVANIA STATE UNIVERSITY

JONATHAN PARSONS

DUKE UNIVERSITY

KATHLEEN M. GATES AND JOSEPH B. HOPFINGER

THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

Significant heterogeneity in network structures reflecting individuals' dynamic processes can exist within subgroups of people (e.g., diagnostic category, gender). This makes it difficult to make inferences regarding these predefined subgroups. For this reason, researchers sometimes wish to identify subsets of individuals who have similarities in their dynamic processes regardless of any predefined category. This requires unsupervised classification of individuals based on similarities in their dynamic processes, or equivalently, in this case, similarities in their network structures of edges. The present paper tests a recently developed algorithm, S-GIMME, that takes into account heterogeneity across individuals with the aim of providing subgroup membership and precise information about the specific network structures that differentiate subgroups. The algorithm has previously provided robust and accurate classification when evaluated with large-scale simulation studies but has not yet been validated on empirical data. Here, we investigate S-GIMME's ability to differentiate, in a purely data-driven manner, between brain states explicitly induced through different tasks in a new fMRI dataset. The results provide new evidence that the algorithm was able to resolve, in an unsupervised data-driven manner, the differences between different active brain states in empirical fMRI data to segregate individuals and arrive at subgroup-specific network structures of edges. The ability to arrive at subgroups that correspond to empirically designed fMRI task conditions, with no biasing or priors, suggests this data-driven approach can be a powerful addition to existing methods for unsupervised classification of individuals based on their dynamic processes.

Key words: multivariate time series, subgrouping, cluster analysis, fMRI, idiographic, individual-level.

1. Introduction

Network-based approaches help researchers understand individual differences in dynamic processes. Here, the variables (or nodes) are observed numerous times for each individual, and the edges between the nodes represent the relations among nodes across time. Oftentimes, the goal is to compare *a priori* subgroups of individuals or see how the structure of a network relates to some correlate of interest. As an example, graph theoretic measures have been used to quantify aspects of functional brain connectivity using summary values to assess between-group differences and correlates (Rosenberg et al., 2016, Nichols et al., 2014). Network measures have also been used in ecological momentary assessment research to identify correlates with static characteristics (Bringmann et al., 2016). The measures often aim to describe the network as a

We acknowledge the primary funding source for this project: National Institute of Health-National Institute of Biomedical Imaging and Bioengineering (R01-EB02290).

Correspondence should be made to Zachary F. Fisher, Quantitative Developmental Systems Methodology Core, Department of Human Development and Family Studies, The Pennsylvania State University, Health and Human Development Building, University Park, PA16802, USA. Email: fisherz@psu.edu

whole (i.e., a summary of all relations or edges among the variables or nodes), such as measures of global efficiency or weight (Rubinov & Sporns, 2010).

These approaches have been helpful in guiding researchers to better understand global features of brain processes that relate to constructs of interest. However, two individuals with the same global measure can have vastly different network structures. Therefore, examining individual-level network structures to identify group differences is also critically important. Such studies have identified specific edges that may be critical or important for a given individual's dynamic process (Epskamp et al., 2018, Fisher & Boswell, 2016). Another important element that is often overlooked in work that examines differences between predefined groups is that heterogeneity may exist within those subgroups. That is, individuals may have more in common with those outside their predefined subgroup than with those inside the subgroup. For these reasons, researchers turn towards unsupervised classification based on the entire network structure.

One approach for identifying individual-level network structures with time series data, GIMME (Group Iterative Multiple Model Estimation; Gates and Molenaar, 2012), complements the use of global measures by providing information regarding how the individuals differ in their network structures. This is done by examining the presence and absence of edges in the network across individuals to find which ones are consistently found for all individuals while also finding nuances in the individual-level pattern. Much work has demonstrated that meaningful within-group heterogeneity in the patterns and weights of edges exists in both clinical (Gates et al., 2014; Yang et al., 2014a; Volkmar et al., 2004) and typically developing control samples (Beltz et al., 2013; Miller & Van Horn, 2007; Scherf et al., 2007), supporting the need for methods that can accommodate the heterogeneity in network structures across individuals.

At present, it is clear that some aspects of dynamic human processes can be expected to vary across individuals (Finn & Constable, 2022, Laumann et al., 2015), while others such as visual processing in the brain may operate more consistently across individuals (Xu et al., 2014). While meaningful heterogeneity may exist across individuals in any given population, there may be subsets of individuals that share similarities in network structures. These commonalities may or may not exist within arbitrary *a priori* subgroup identifications (such as diagnostic category or gender) as these themselves may be heterogeneous. Hence, methods are needed for identifying clusters of individuals who share similar network structures that represent individuals' dynamic processes. The recently developed subgrouping-extension within GIMME (S-GIMME) was designed to enable researchers to identify meaningful clusters of individuals (Gates et al., 2017). Critically, S-GIMME provides subgroup assignments of individuals based entirely on network structures of their dynamic process, as opposed to subjective measures (e.g., diagnoses).

S-GIMME differs from other personalized approaches for arriving at sparse network structures in a number of ways. First, the classification of individuals is unsupervised rather than supervised. Whereas network structures have been used as input to identify categories of individuals from known groups (e.g., diagnoses; Dickie et al., 2018), S-GIMME does not predict or classify individuals into *a priori* groups. Hence, this is not an approach where a portion of the individuals are used to train and another portion to test, as we are not conducting supervised learning. Second, it does not aim to predict scores on other constructs of interest (e.g., Dubois et al., 2018; Shirer et al., 2012); the purpose of S-GIMME is to identify subsets of individuals who share similarities in their network structures.

Extensive simulation studies have evaluated the use of S-GIMME (Gates et al., 2017, Lane et al., 2019). The general findings are that with enough time points ($t \geq 60$) and individuals ($k \geq 25$), S-GIMME can identify the subset of individuals whose data were generated from similar network structures. It also can recover the network structures from the generated time series data. While S-GIMME has been used on empirical data both in functional MRI (e.g., Price et al., 2017a; Dajani et al., 2019) as well as ecological momentary assessment data (e.g., Wright et al., 2019), it has yet to be evaluated using an experimental design to ensure the subgroups relate

to meaningful constructs outside the system of dynamic variables in the model. Here, we validate the use of S-GIMME using a tightly controlled experimental design. Functional MRI data were collected with the purpose of evaluating the ability to correctly separate individuals' data sets based solely on differences in their dynamic brain processes while engaged in different tasks. We first describe some background information on the use of networks with functional brain data before presenting the details of the study and results.

2. Subgrouping GIMME

S-GIMME analyses were conducted using the `gimme` R package (Lane et al., 2021, https://tarheels.live/gimme/). Although the GIMME and S-GIMME algorithms have been explained in detail elsewhere (see Beltz & Gates, 2017; Gates & Molenaar, 2012; Lane & Gates, 2017), in the following section we provide a concise technical review of the model and algorithms underlying S-GIMME, for completeness.

2.1. The Structural VAR Model

Formally, GIMME and S-GIMME operate on a multivariate time series composed of d component series, $\{\mathbf{X}_t\}_{t\in\mathbb{Z}} = \{(X_{j,t})_{j=1,\ldots,d}\}_{t\in\mathbb{Z}}$, where $\mathbf{X}_t$ follows a structural vector autoregressive model of order 1 (SVAR; Lütkepohl, 2005, p. 358),

$$\mathbf{X}_t = \boldsymbol{A}\mathbf{X}_t + \boldsymbol{\Phi}\mathbf{X}_{t-1} + \mathbf{E}_t, \quad t \in \mathbb{Z}, \tag{1}$$

where $\mathbf{A}$ contains the contemporaneous directed relations among $\mathbf{X}_t$, $\boldsymbol{\Phi}$ contains the auto- and cross-regressive relations between $\mathbf{X}_t$ and $\mathbf{X}_{t-1}$, and $\mathbf{E}_t$ are process noise with a strictly diagonal covariance matrix, $\boldsymbol{\Xi} = \boldsymbol{\Sigma}_{\mathbf{E}}$. In the presentation that follows, we only consider weakly stationary series for a linear SVAR model.

A typical approach for fitting such a model involves first fitting a canonical vector autoregressive model (VAR; Lütkepohl, 2005, p. 13) of identical order, and recasting the coefficients by means of a Cholesky decomposition of the full covariance matrix of the process innovations (see Molenaar, 2017, p. 252 for details of this transformation). Unfortunately, SVAR coefficients estimated by means of Cholesky decomposition are not unique; solutions depend on the arbitrary ordering of the univariate components in $\mathbf{X}_t$. In addition to the explicit goal of identifying generalizable and person-specific dynamics, the GIMME algorithm provides a data-driven approach for uniquely identifying the SVAR model above. This will be discussed in greater detail within the context of S-GIMME below.

2.2. Group, Subgroup and Individual-Level Models

To begin our discussion of the S-GIMME algorithm, consider a zero-mean multivariate time series $\mathbf{X}_t^k$ for a single individual k,

$$\mathbf{X}_t^k = \boldsymbol{A}^k\mathbf{X}_t^k + \boldsymbol{\Phi}^k\mathbf{X}_{t-1}^k + \mathbf{E}_t^k, \quad t \in \mathbb{Z},\ k = 1, \ldots K \tag{2}$$

for some $d \times d$ matrix of contemporaneous directed relations $\boldsymbol{A}^k$, a $d \times d$ matrix of lagged directed relations $\boldsymbol{\Phi}^k$, and a white noise series $\{\mathbf{E}_t^k\}_{t\in\mathbb{Z}} \sim \text{WN}(\mathbf{0}, \boldsymbol{\Xi}^k)$ characterized by $\mathbb{E}(\mathbf{E}_t^k) = 0$ and $\mathbb{E}(\mathbf{E}_t^k\mathbf{E}_{s'}^k) = 0$ for $s \neq t$. Importantly, in the current presentation, $\boldsymbol{\Xi}^k$ is strictly diagonal (though see Luo et al., 2022, for details on relaxing this assumption).

The S-GIMME framework relies on the following decomposition of the weight matrices $\boldsymbol{A}^k$ and $\boldsymbol{\Phi}^k$ into group, subgroup and individual components

$$\boldsymbol{A}^k = \boldsymbol{A}_g^k + \boldsymbol{A}_s^k + \boldsymbol{A}_k^k \tag{3}$$

$$\boldsymbol{\Phi}^k = \boldsymbol{\Phi}_g^k + \boldsymbol{\Phi}_s^k + \boldsymbol{\Phi}_k^k \tag{4}$$

where $\boldsymbol{A}_g^k$, $\boldsymbol{\Phi}_g^k$, $g = G = 1$, are the group-level contemporaneous and lagged weight matrices, $\boldsymbol{A}_s^k$, $\boldsymbol{\Phi}_s^k$, $s = 1, \ldots, S$, are the subgroup-level contemporaneous and lagged weight matrices, and $\boldsymbol{A}_k^k$, $\boldsymbol{\Phi}_k^k$, $k = 1, \ldots, K$ are the individual-level contemporaneous and lagged weight matrices, respectively.

Broadly, the subscripts g, s and k indicate whether the pattern of relations (location of zero and nonzero elements) is held constant at the group, subgroup or individual level, respectively. The superscript k is used to indicate individual-level variability in the weights associated with a given subscript value. For example, the group-level weight matrices, $\boldsymbol{A}_g^k$ and $\boldsymbol{\Phi}_g^k$, $g = 1$, will have the same pattern of zero and nonzero elements across all individuals in the sample, while the weights associated with the nonzero elements can vary across the $k = 1, \ldots, K$ individuals. Similarly, the subgroup-specific weight matrices, $\boldsymbol{A}_s^k$ and $\boldsymbol{\Phi}_s^k$, $s = 1, \ldots, S$, will have the same pattern of zero and nonzero elements for all individuals within a given subgroup, while the weights associated with those nonzero elements can vary across subgroup members. Lastly, the individual-specific weight matrices, $\boldsymbol{A}_k^k$ and $\boldsymbol{\Phi}_k^k$ are allowed to vary both in their pattern of relations and weights, across all individuals.

2.3. The S-GIMME Algorithm

The group, subgroup and individual-level estimates outlined above are obtained across three sequential stages of the S-GIMME algorithm: *group stage*, *subgroup stage*, and *individual stage*.

The Group Stage The S-GIMME algorithm begins by estimating a constrained version of (1) for each individual separately, where only the diagonal elements of $\boldsymbol{\Phi}^k$ and $\boldsymbol{\Xi}^k$ are freely estimated. Beginning the model search with autoregressive paths freely estimated (and not constrained to zero) brings a number of practical and theoretical benefits, which have been discussed in detail elsewhere (Beltz & Molenaar, 2016; Luo et al., 2022; Weigard et al., 2021). Here, we mention two important considerations with respect to the analysis of fMRI data. First, the time course of the BOLD signal typically results in a persistent signal with strong autocorrelations among the component series (Olszowy et al., 2019). Second, the inclusion of the autoregressive parameters helps to identify the directionality of off-diagonal elements of $\boldsymbol{A}^k$. Without this (or some similar identification constraint), it is not possible to identify the directed contemporaneous paths between two component series (see Beltz and Molenaar, 2016, for alternative approaches to identification issues in the SVAR model).

With each individual's baseline model estimated, modification indices (MI; Sörbom, 1989) are obtained for each fixed-to-zero element in $\boldsymbol{A}^k$ and $\boldsymbol{\Phi}^k$. These MIs indicate which relation, if freed, would improve a given individual's model fit by explaining additional variance in the outcome node. Modification indices are asymptotically χ^2 distributed, enabling significance testing. Internally, GIMME records the proportion of individuals a given MI is significant following a strict Bonferroni correction. Edges (or elements in $\boldsymbol{A}^k$ and $\boldsymbol{\Phi}^k$) are then iteratively added to the *group-level* structure (i.e., the pattern of edges that will be estimated for all individuals in the sample) if they are both (1) significant for a predefined proportion of individuals, and (2) have the highest count (proportion of individuals with a significant MI) relative to other edges. This continues until no edge is significant for the majority of the sample, where the proportion needed to make it a group-level path is defined by the researcher.

By default, 75% of the sample must have an edge for it to be considered a group-level path. Considering the signal-to-noise ratio and time series length for a given set of data, there may not be sufficient power to detect relations for the entire sample; therefore, a value lower than 100% is preferred. By counting the number of significant modification indices rather than summing or averaging them, this method is also more robust to outlier cases. In this way, group-level inferences are more accurate than approaches that first aggregate the data (e.g., concatenating time series or averaging correlation matrices) as spurious edges are not included in any other individual's model. At the conclusion of the group stage, the pattern of edges in $\boldsymbol{A}_g^k$ and $\boldsymbol{\Phi}_g^k$ is established for all individuals in the sample. These edges are freely estimated in all subsequent stages of the analysis.

The Subgroup Stage The *subgroup stage* of S-GIMME begins by constructing a $K \times K$ similarity matrix based on the group-level results obtained in the *group stage*. This similarity matrix is constructed in the following manner. For each individual, the sign and significance of each edge in $\boldsymbol{A}^k$ and $\boldsymbol{\Phi}^k$ are determined. If the edge exists in the group-level model, sign and significance are obtained from the estimated coefficient, associated standard error, and Bonferroni-corrected p-value. If the edge is not freely estimated in the group-level model, the expected parameter change (EPC; Saris et al., 1987) is used to determine the edge's sign and Bonferroni-corrected significance. With the sign and significance of each element in $\boldsymbol{A}^k$ and $\boldsymbol{\Phi}^k$ available, a similarity matrix is constructed where each element represents a count of the number of edges that are similar among two given individuals in terms of being significant and of the same sign. Finally, sparsity is then induced by subtracting the lowest value in the matrix from all edges.

The community detection algorithm Walktrap (Pons & Latapy, 2005) is then applied to the similarity matrix with the aim of clustering individuals based on their network structures(Gates et al., 2017). Walktrap has been found to perform well on networks with a range of characteristics, including count and correlation matrices as well as small and large networks (Gates et al., 2016; Golino & Epskamp, 2017). The intuition behind Walktrap rests on the idea that one can estimate the probability two nodes (or communities) are reachable using random walks of a given length on a graph (e.g., a similarity matrix). These random walks tend to reveal densely connected areas representing communities; the distance metric used by Walktrap is based on this idea of random walks, and the probability two nodes (or communities) are connected. Walktrap proceeds by performing short random walks, typically three to five steps, and then merges communities using Ward's method (Ward Jr & Hook, 1963) and modularity (Newman, 2004) to cut the dendrogram (determine the optimal number of clusters). Although it is beyond the scope of this paper, Pons & Latapy (2005) provide formulas and a technical development of the Walktrap algorithm.

Once subgroup labels are obtained, a search for paths that exist for the predefined majority of individuals within each subgroup is conducted following the procedure described for the group-level paths. Here, a 51% or 75% cut-off for what constitutes the majority is typically used to account for the smaller sample size within subgroups. Importantly, the proportion of individuals chosen to constitute the majority (e.g., 51% or 75%) does not impact the subgroup assignments. At the conclusion of the subgroup stage, the pattern of edges in the group-level ($\boldsymbol{A}_g^k$, $\boldsymbol{\Phi}_g^k$) and subgroup-level matrices ($\boldsymbol{A}_s^k$, $\boldsymbol{\Phi}_s^k$, $s = 1, .., S$) are established. These edges are then freely estimated for each individual separately during the final, individual stage.

The Individual Stage Using the estimates obtained from the group- and subgroup-level network structures as a foundation, the algorithm searches for additional lagged and contemporaneous edges that are needed to best explain each individual's dynamic process. Fit indices are used as stopping criteria, with the search for individual-level paths ending once two out of four of the following are found to be acceptable using predefined cutoffs: Tucker-Lewis Index (TLI), Non-normed Fit Index (NNFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean-Squared Residual (SRMR). Values greater than 0.95 for the TLI

and NNFI indicate an acceptable fit and values lower than 0.05 for RMSEA and SRMR are considered acceptable (Brown, 2006). This approach favors parsimony in the pattern and prevents the emergence of false positives. In the end, the researcher is provided with labels for which subgroup each individual belongs to, as well as individual-level estimates of dynamic relations (both contemporaneous and lagged) and whether they are a group, subgroup or individual-level edge.

3. Networks in fMRI

Researchers conceptualize mental processes as being the coordinated activity of spatially disparate brain regions across time (Bullmore & Sporns, 2009; Sporns, 2016). This represents a paradigm shift from localizing isolated brain regions that independently relate to a given condition. With this shift came an influx of new types of analyses, with network theory perspectives prevailing. Viewing the brain from a network perspective is relatively new, and researchers often do not have all the information necessary to build concrete hypotheses regarding how brain processes differ meaningfully across individuals and conditions. For this reason, data-driven approaches for arriving at network structures that depict brain processes are especially critical. In this way, the field of brain sciences is not so dissimilar from the fields encompassing social and psychological sciences as oftentimes there are not clear hypotheses regarding how individuals may differ in their dynamic processes in these fields too.

3.1. Subgrouping on Network Dynamics in fMRI

Perhaps unsurprisingly, a number of (overlapping) taxonomies have been used to delineate subgrouping algorithms in terms of their underlying methodologies and objectives. Broad surveys of different approaches to time series clustering or subgrouping are available in Liao (2005) and Aghabozorgi et al., (2015). In the context of fMRI analyses, a number of important distinctions related to subgrouping procedures can be made. First and foremost, a distinction can be made between approaches that cluster individuals into meaningful clusters or groups, and the more common application of clustering voxels or pre-defined regions of interest (ROIs) into functionally similar groupings. The latter is typically concerned with dimension reduction for the purpose of identifying discrete functional units (voxels or ROIs) with similar temporal signal characteristics (Heller et al., 2006) or to identify subnetworks of ROIs that tend to covary across time (e.g., the default mode network; Power et al., 2011). Although an important use case for clustering methods in fMRI, these approaches are not directly relevant to the current work.

Additional distinctions can be made within approaches designed to subgroup individuals. For example, modern clustering methods are commonly separated into two classes: nonprobabilistic and probabilistic approaches (Ernst et al., 2021). Nonprobabilistic approaches, such as k-means clustering, typically use heuristic functions to minimize a criterion that produces nonprobabilistic subgroup assignments. On the other hand, probabilistic approaches typically utilize a statistical model, such as a Gaussian mixture model, to assign each individual a probability of belonging to a given subgroup or cluster. Although historically less common, recent interest in disease subtyping has led to an uptick in papers using unsupervised classification methods on fMRI data (Miranda et al., 2021). Common to many of these approach is the use of connectivity weights to reduce the dimension of the classification problem relative to the raw data, followed by the application of probabilistic and nonprobabilistic subgrouping approaches to identify similar patterns of dependence among individual dynamics (Yang et al., 2014b; Brodersen et al., 2014; Tokuda et al., 2018).

Similarities between these methods and S-GIMME exist. Namely, S-GIMME could be considered a nonprobabilistic approach using connectivity weights during subgrouping. A number of

important features make S-GIMME unique among other approaches commonly used with fMRI (and multivariate time series from multiple subjects more generally). In their review of methods for dynamic clustering, Ernst et al. (2021) note S-GIMME is exceptional among other approaches in that it estimates relations at the group, subgroup and individual-level, while the majority of approaches estimate relations entirely at the subgroup-level. In addition to helping characterizing dynamics at important levels of the analysis, this places fewer assumptions on the homogeneity of individual dynamics, at the individual or subgroup level.

Although not highlighted in the current work, there are additional capabilities that make S-GIMME an attractive approach for those interested in subgrouping fMRI data. S-GIMME can handle missing data using Full-Information Maximum Likelihood (FIML; Enders, 2001). The modeling of direct and modulatory effects of tasks using person-specific finite impulse response function (hemodynamic response function; Duffy et al., 2021) and exogenous variables (Arizmendi et al., 2021) is easily accommodated within the subgrouping procedures. When subgroups or clusters are known *a priori* and the interest is in better characterizing the heterogeneity in network dynamics among these groups, confirmatory options are available (Henry et al. 2019a). Lastly, it is also possible to analyze between-network relations and higher-dimensional fMRI data using latent variables and principal components options (Gates et al., 2020).

3.2. Empirical Evaluation

Since behavioral and emotional studies under the ecological momentary assessment framework often take a matter of weeks or months with reactions to an experimental manipulation unknown, it was more realistic to design a controlled experiment for validation that occurs on functional brain data where the expected differences in brain networks between tasks are already well-established.

We thus evaluated the ability of the S-GIMME algorithm to provide accurate data-driven identification of brain networks and to subgroup individuals using empirical task-based fMRI data. While recent work has shown the ability of S-GIMME to recover subgroup assignments and network structures in simulated data (Gates et al., 2017; Lane et al., 2019), never before has the ability to recover expected results been tested in a controlled experimental design. In the current study, we developed 3 fMRI task designs to evoke distinct network structures. The aim was to evaluate if the algorithm would differentiate between the conditions, in a completely blind, data-driven analysis. To increase the difficulty of this classification, each subject participated in each of the three conditions. This allowed us to ascertain if the analyses could subgroup accurately based on the task being performed rather than simply clustering individuals with their own other runs. With no *a priori* information regarding condition, behavior, or individual characteristics, the unsupervised approach was tested to examine how well it could separate the data by task using only the fMRI time courses. These results should be applicable to research across varied domains, as the methods rely on purely bottom-up clustering of individuals based on time series. This is also relevant to studies that compare different populations or search for possible subgroups within a population (e.g., diagnostic status as in Price et al., 2017b).

We used fMRI data from a passive visual processing task versus an auditory-motor task (with no visual component) versus a task that includes both of these processes plus the addition of a higher-order visual process (face-perception). During the tasks that require an overt motor response, the coefficient between the motor regions and the sensory areas should be stronger. During the task with faces, the edges with the fusiform face area (FFA; Kanwisher et al., 1997) should be evident. Here, we test whether S-GIMME is sensitive enough to be able to group data by task, even when the difference between tasks is relatively subtle (e.g., scrambled versus intact faces during visual tasks, and eyes-open versus eyes-closed during motor tasks) and largely restricted to edges with just one or two nodes of the network.

To contextualize the performance of S-GIMME, we also tested whether two alternate time series clustering approaches could differentiate between the different task conditions. To overcome the curse of dimensionality ((Bellman, 1966) inherent to clustering multivariate time series from multiple individuals, individual time series were characterized by their VAR(1) transition matrices. After reducing the individual time series to the individual-level dynamics, we considered a probabilistic and nonprobabilistic approach. The probabilistic approach used a Gaussian finite mixture model (McLachlan & Chang, 2004) to cluster and the Bayesian information criterion (BIC; Schwarz, 1978) to select the number of subgroups. For the nonprobabilistic approach, we used the popular k-means algorithm and the Calinski–Harabasz index (Caliński& Harabasz, 1974) for estimating the number of subgroups. Both the k-means algorithm (Easson et al., 2019; Gumus et al., 2021) and Gaussian mixture modeling (Brodersen et al., 2014; Gumus et al., 2022) have been successfully employed in applications designed to subtype individuals using fMRI data.

4. Methods

4.1. Participants

Participants were screened to ensure that they were free of neurological or psychiatric disorders, had normal or corrected-to-normal vision and were right-handed. Thirty-three healthy young adults enrolled in the study and completed all runs successfully; each was paid $20/hour. Three subjects were excluded for excessive motion during the fMRI runs. The data reported here are from the remaining thirty participants (ages 18–28, mean age 22.8, 18 female). All procedures were approved by the Institutional Review Board at the University of North Carolina at Chapel Hill, and all participants provided written informed consent.

4.2. Tasks/fMRI Runs

Participants viewed images on a translucent screen through an angled mirror attached to the head coil. OpenSesame software (https://osdoc.cogsci.nl; Mathôt et al., 2012) was used to present stimuli and record button-press responses. Each participant completed three epoch-design (block-design) runs. Each run was 448 s in total duration and is described below. After these runs, subjects performed three additional runs that are not reported here. Prior to beginning the MRI session, participants completed a brief (2-min) training outside of the scanner, to ensure subjects understood the tasks.

4.2.1. Visual-Only Condition This condition was designed to engage visual processing, without linking this to higher-order face processing or to the motor or auditory systems. The sequence of events is illustrated in Fig. 1 (left column). The images were scrambled versions of the face stimuli used in the face-motor-auditory run described below. The images were scrambled by dividing the face images into 20×20 pixel squares and randomly shuffling the positions of those squares. Each image was presented for 500 msec, followed by 300 msec of a fixation-cross-only screen. This is repeated (with different scrambled images) 20 times for each 16-s visual-stimulation block. Following each 16-s visual-stimulation block was a 16-s fixation-only block, during which participants were asked to keep their eyes fixated on the central cross throughout the block. Participants were instructed to passively view the stimuli; no button press responses were required at any time. This block design (visual stimuli for 16 s, followed by fixation-only for 16 s) was repeated 14 times, resulting in a total duration of 448 s. This was the first fMRI run for each subject.

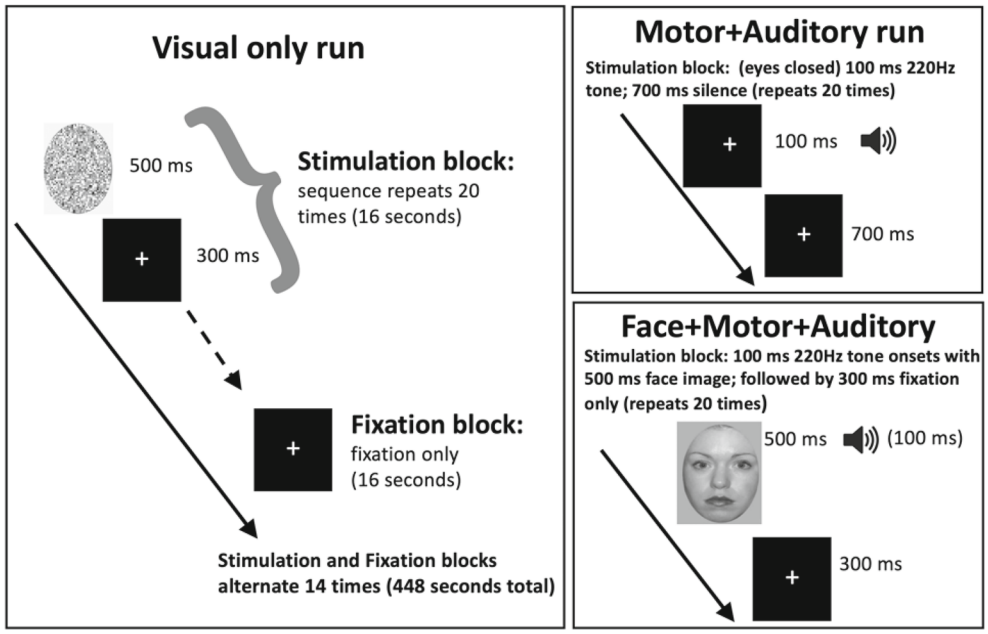

FIGURE 1.
Trial sequence and stimuli for the 3 task runs. All 3 runs consisted of alternating 16 s of Stimulation blocks and 16 s of Fixation/Silence Blocks (in the case of the Motor-Auditory run, this was simply a quiet no-stimulus block that alternated with the blocks of tones+motor responses). In the Face-Motor-Auditory run, the tone and face onset simultaneously, with the durations being 100 and 500 msec, respectively. The 16 s fixation-only (no-stimulus) block that alternated with that task blocks is only shown for the Visual-only condition, but it occurred in all task runs.

4.2.2. Motor-Auditory Condition This condition was designed to engage the motor and auditory systems, without engaging the visual processing systems. In this condition, participants were instructed to keep their eyes closed for the entire duration of the run, and to press the button on the MRI-compatible response box (Current Designs Pyka 5-Finger Response Pad) every time a specified tone was played through the headphones. The tone was presented at the same rate (once every 800 msec) as the scrambled-face image was presented in the Visual-only bock. This was done to ensure that the motor system was engaged at the same temporal frequency in this condition as the visual system was engaged in the Visual-Only condition. Since the trigger to make the behavioral response was an auditory tone, this could result in a linking of the motor and auditory systems, even though overall activity in the auditory system could be at a high level already due to the significant auditory noise of fMRI scanning. The sequence of events is illustrated in Fig. 1 (top right). The auditory stimulus was a 220 Hz tone, played for 100 msec. There was 700 msec between successive auditory tones, and participants were asked to press the response button after each tone. This was repeated 20 times for each 16-s response-block. Following each 16-s response-block was a 16-s block with no stimulation or responses. This sequence repeated 14 times, for a total duration of 448 s. This condition was always completed after the Visual-only run.

4.2.3. Face-Motor-Auditory Condition This condition was designed to engage all of the systems engaged in the Visual-only and the Motor+Auditory conditions, plus the fusiform face area. Unlike the Visual-only condition, this condition was expected to show a linkage between visual and motor regions. The highly similar response requirements between the Motor-Auditory and Face-Motor-Auditory conditions, with the addition of face processing in the latter, provide a test

of whether our software is sensitive enough to detect these relatively small differences between conditions, even if the same subjects are performing all tasks. The visual images were faces taken from the NimStim database (Tottenham et al., 2009); 20 face images were used (each appeared once in every 16-s stimulation period, in randomized order), and the auditory stimulus (220 Hz tone) was the one used in the Motor-Auditory condition. The sequence of events is illustrated in Fig. 1 (right bottom). On each trial, a face image was presented simultaneously with the tone. The tone was presented for a duration of 100 msec, and the face image was presented for 500 msec. Subjects were instructed to press the response button for each presentation of the face/tone. Following the offset of the face, there was 300 msec of a fixation-cross-only screen. This sequence repeated (with different faces) 20 times for each 16-s stimulation block. Following each 16-s visual-stimulation block was a 16-s fixation-only block. This block design repeated 14 times, for a total duration of 448 s.

4.3. MRI Scanning Protocol

Images were collected on a 3T Siemens PRISMA MRI system at the University of North Carolina Biomedical Research Imaging Center. Functional images included 37 transverse slices ($3 \times 3 \times 3\ mm^3$ resolution), collected interleaved inferior to superior. Images were acquired using a T2-weighted echo-planar imaging (EPI) sequence (TR = 2000 ms, TE = 26 ms, flip angle = 80°). After discarding the first three scans to allow for magnetic field stabilization, the fMRI runs reported here each lasted 448 s. A structural scan was also acquired for each participant (T1-weighted; TR = 2400 ms; TE = 2.22 ms; flip angle = 8, FOV = 256 mm, 208 slices, $0.8 \times 0.8 \times 0.8\ mm^3$ resolution; 398 s duration).

4.4. fMRI Processing

MRI data were preprocessed using SPM12 (Wellcome Department of Imaging Neuroscience, University College London, UK). Preprocessing included spatial realignment and slice-time correction. The mean image constructed from realignment was used to determine parameters for coregistration and spatial normalization into the standard MNI-space using the EPI-template included in SPM12. The fMRI data were normalized at a $2 \times 2 \times 2$ mm resolution and were then smoothed with an 8 mm full-width at half-maximum isotropic kernel. For the analysis of task-dependent activation, covariates of interest specified each trial in terms of stimulus onset (e.g., scrambled-face onset; auditory tone onset; face-onset). Six movement parameters obtained during realignment were included as covariates of no interest for each run, and an additional covariate for each run incorporated an overall intercept to the model. Each regressor was convolved with canonical response hemodynamic function and entered into the general linear model (Friston et al., 1994). Contrast maps for each subject were entered into a second-level analysis.

4.5. Regions of Interest

Brain regions of interest (ROI) were chosen to cover both regions thought to be involved in the conditions being compared and regions thought to be engaged independently from our tasks (those in the default mode network, DMN). This allowed us to examine whether S-GIMME results would be driven primarily by edges corresponding to regions related to the tasks versus possibly spurious differences across individuals in an independent network. Also, in contrast to graph network summary methods that utilize hundreds of ROIs, the GIMME algorithm is intended to investigate directed edges (as opposed to correlational or undirected relations) across a limited number of regions. Thus, we chose 2 primary visual regions, 2 primary auditory regions, 2 motor

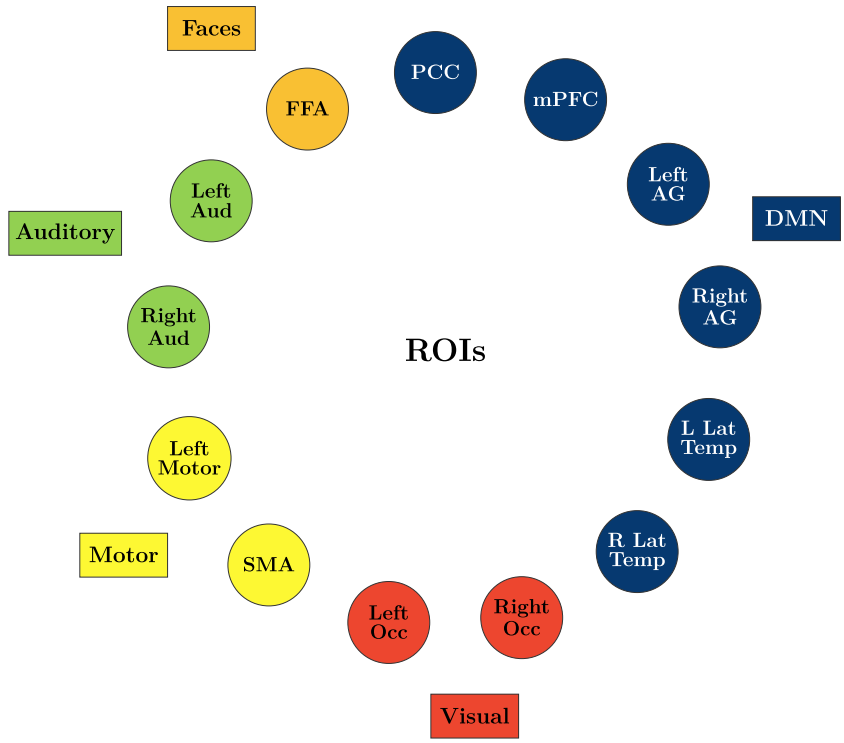

FIGURE 2.
Regions on Interest (ROIs) for all analyses, laid out in positions used in all subsequent analysis plots. Abbreviations labeled on ROIs are defined as in Table 1. DMN=Default Mode Network.

regions, the face-selective fusiform face area, and 6 regions of the DMN, distributed across anterior and posterior regions and hemispheres. For the DMN, we chose the 6 regions defined and labeled in Power et al., (2011) supplemental materials for their meta-analytic analyses. The task-based regions were defined based on the average location of the peak coordinate from the individual subjects' fMRI results (as described below) within each of the general areas described above (e.g., primary visual cortex, etc.), using Neurosynth (https://neurosynth.org) to define general regions. A voxel of peak activation within each region was identified for each participant individually. The coordinates of the voxels for each region were then averaged across participants, and a 10 mm diameter sphere was built around the average coordinate, which was then used as the ROI for that region for every participant.

The regions, networks, and center coordinates of all of the ROIs are listed in Table 1 and shown graphically in Fig. 2. MarsBaR software (Brett et al., 2002; https://sourceforge.net/projects/marsbar/) was used to generate the ROIs and extract the BOLD time-series for each ROI, which was defined as the average of time courses for all voxels within each 10 mm ROI sphere. We also extracted regions of white matter (WM) and cerebral spinal fluid (CSF), defined for each participant based on the tissue probability maps from SPM segmentation, to regress out the contributions from physiological artifacts. Time series data from the WM and CSF ROIs were used as nuisance regressors when extracting time series data from the ROIs for the main analyses.

TABLE 1.
ROI coordinates in MNI (montreal neurological institute) space. Left and Right refer to brain hemisphere.

Network	Region	x	y	z
Visual	Left Occipital (primary visual)	−12	−96	4
Visual	Right Occipital (primary visual)	16	−96	10
Motor	Left Motor (primary; M1)	−44	−12	48
Motor	Supplementary Motor Area (SMA)	−4	−2	56
Auditory	Left Auditory (primary)	−58	−20	14
Auditory	Right auditory (primary)	60	−12	12
Face processing	Fusiform Face Area (FFA)	36	−62	−18
Default mode network	Posterior Cingulate Cortex (PCC)	1	−51	29
Default mode network	Medial Prefrontal Cortex (mPFC)	−1	61	22
Default mode network	Left Angular Gyrus (AG)	−48	−66	34
Default mode network	Right Angular Gyrus (AG)	53	−61	35
Default mode network	Left Lateral Temporal (lat temp)	−65	−23	−9
Default mode network	Right lateral temporal (lat temp)	61	−21	−12

5. Results

5.1. Standard fMRI Results

For each task, we conducted epoch-based random effects analyses, comparing task blocks to fixation blocks, to ensure our tasks were evoking signal in the expected regions. Visual regions were highly active in the Visual-only condition, the Motor+Auditory condition revealed robust activity in motor and auditory regions, and the Face-Motor-Auditory contained all those regions, plus additional activity in the region of the FFA (Fig. 3).

5.2. Network Results

5.2.1. Each Condition Alone We first analyzed each of the 3 task conditions separately, to see if GIMME would find network structures that aligned with the expectations of what each task alone should generate, without finding spurious edges. As shown in Fig. 4, the algorithm was effective in doing this. The Visual-only condition was found to have persistent edges between the primary visual areas, and these were not robustly linked to the auditory, motor, or face areas. In addition, while the DMN areas were robustly connected with each other, none of these were linked to the visual processing regions. The Motor-Auditory condition was similar to the Visual-only condition, with the critical addition of a consistent link between the motor and auditory regions, and another consistent link between the 2 motor regions. These additions align with the requirement that subjects produce button presses in response to the auditory tone. Finally, the Face-Motor-Auditory condition result was highly similar, with the addition of a consistent edge between the visual areas and the FFA, in line with this being the only condition with face stimuli. These analyses provide evidence that the original GIMME algorithm is able to capture, in a purely bottom-up and data-driven manner, the network structures, within and across networks, expected to be seen within each of these conditions. Importantly, each condition was analyzed the same, with no biasing and no priors based on expected results; yet the algorithm found network structures that differed across tasks, specifically in the edges expected to be different.

To better understand the homogeneity within each task condition, we also ran S-GIMME on each task condition alone. For all tasks S-GIMME identified two subgroups. For the Visual-only condition, one subgroup resembled the group model, while the other subgroup included

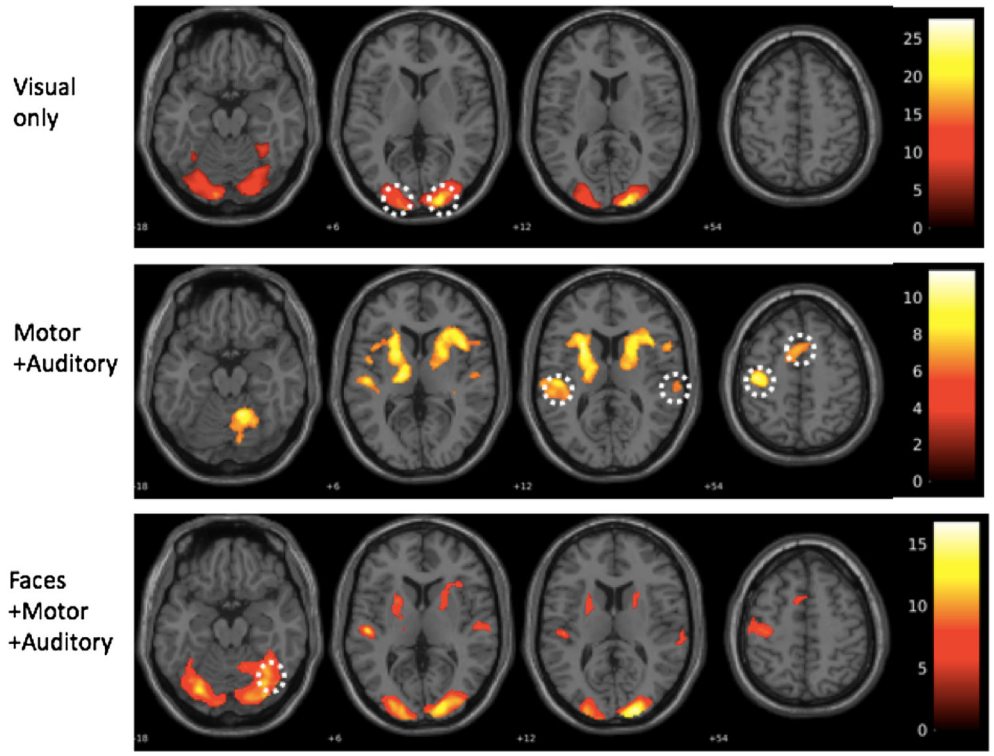

FIGURE 3.
Random effects analysis results for all 3 tasks, analyzed separately, comparing task blocks to fixation blocks (all maps shown at $p < .05$ FWE-corrected). Dashed white circles indicate approximate location of ROIs, with Primary Visual ROIs drawn on Visual-only task, Primary Auditory and Motor ROIs drawn on Motor-Auditory task, and FFA drawn on Face-Motor-Auditory task.

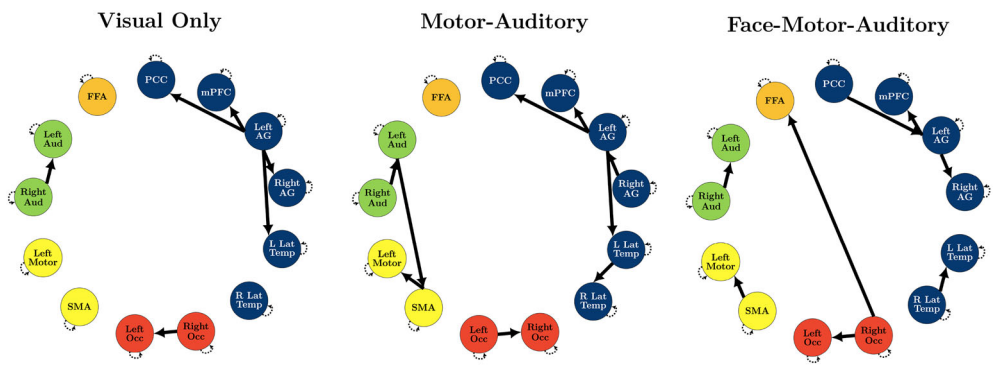

FIGURE 4.
GIMME results from when each task was analyzed separately. Thick black lines represent edges that were consistent across all subjects (significant for at least 75%), separately for each condition. Regions labeled and defined as in Table 1 and Fig. 2.

three additional edges: one additional edge between the DMN regions, one between the motor regions, and one between the motor and auditory areas. Importantly, the main conclusion from the visual-only condition held in that the connection between visual areas was persistent in each subgroup, as expected. For the Motor-Auditory condition, two subgroups were identified; however, no additional subgroup edges were recovered. This can occur when two subgroups are identified, but no additional paths improve fit for the predefined majority of individuals within a subgroup. Finally, the Face-Motor-Auditory condition also identified three additional subgroup edges in one subgroup, and no additional paths in the other subgroup. In addition, both groups showed the persistent connection between visual areas and the face-specific FFA, and none of the additional connections were with the FFA.

5.2.2. All Conditions Together; All Subjects in All Conditions We investigated whether the S-GIMME algorithm could subgroup data when all tasks and all subjects were combined. In the S-GIMME analysis in which all runs from all subjects were included (all 30 subjects contributed 1 run in each of the 3 conditions for a total of 90 datasets), the results produced 2 subgroups, as shown in Fig. 5. Besides one additional subgroup path within the DMN areas, the major difference between these subgroups is that only one subgroup has edges between FFA and visual regions, while the other subgroup has an edge between the auditory and motor regions. The subgrouping corresponds well to two of the conditions, as 28 of the 30 Motor-Auditory runs were placed in the group without a edge between FFA and primary visual regions, whereas 20 out of the 30 Face-Motor-Auditory runs were placed in the subgroup with an edges to the FFA. While S-GIMME separated these two conditions well, the Visual-only condition was split between the 2 subgroups. This split result will be discussed more below, but this may highlight an important point to consider in future studies that aim to use data-driven approaches; specifically, that there may be greater variability during fMRI runs that do not entail a cognitively demanding task (such as the passive viewing of the visual-only condition here).

5.2.3. Paired Conditions In order to provide a direct test of whether the algorithm can separate 2 conditions, we performed 3 sets of analyses with S-GIMME, in which each comparison contained only 2 tasks. It is important to point out that the S-GIMME algorithm is not forced to produce any certain number of subgroups; results can vary from one extreme of producing no subgroups at all to the other extreme of identifying a different subgroup for each subject. Therefore, it is important to point out that although the following analyses each include only 2 conditions, the algorithm had no input that there were 2 conditions; all analyses were conducted in a purely data-driven manner, on the basis of 60 fMRI time-series (i.e., two from each person).

When the 2 tasks were the Motor-Auditory and the Face-Motor-Auditory, the S-GIMME analyses produced 2 subgroups that were very similar in terms of their network structures, with the key differences being that in only 1 of the subgroups there was a strong edges between the FFA and primary visual cortex, and in the other subgroup a strong connection between the auditory and motor areas (Fig. 6). These subgroups corresponded well with the tasks. Specifically, of the 30 Face-Motor-Auditory runs, 27/30 were sorted into the group that contained a strong edge between FFA and primary visual cortex, whereas for the 30 Motor-Auditory runs, 27/30 were sorted into the group without an edge between FFA and visual cortex. This is a 90% classification accuracy in clustering individuals with others who are performing the same task.

The comparison of Motor-Auditory and Visual-only also produced exactly 2 subgroups, as shown in Fig. 7. In this analysis, there was some consistency for the Motor-Auditory condition, as 20/30 runs of this condition were grouped together, with a network that included a persistent edge between Motor and Auditory regions. The Visual-only group also showed some consistency with 24/30 runs being assigned to subgroup 1; however, no subgroup paths emerged for this set of individuals, making the distinction less clear. This is a 73% classification accuracy in clustering individuals with others who are performing the same task.

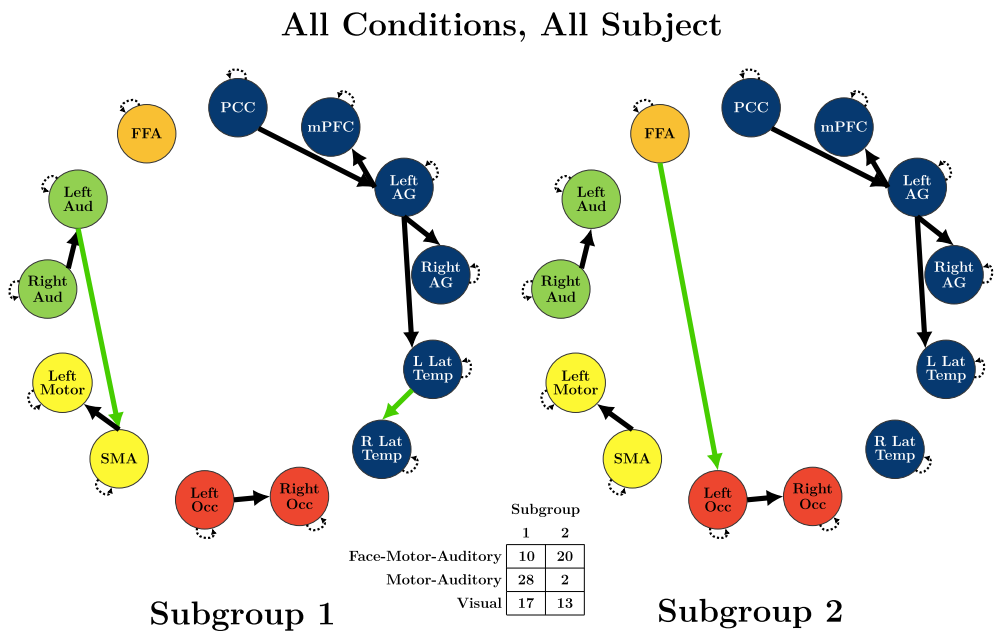

FIGURE 5.
S-GIMME results when all 3 conditions analyzed together, with each subject contributing data to each condition. Thick black lines represent edges that were consistent across all subjects (significant for at least 75%, across all groups). Green lines represent subgroup-level edges (edges that were present in a majority of subjects within a subgroup). Regions labeled and defined as in Table 1 and Fig. 2. The inset table indicates how many runs (out of 30) from each condition were included in that subgroup.

The comparison of Face-Motor-Auditory and Visual-only also produced exactly 2 subgroups (Fig. 8). There was not a strong separation between these two conditions, consistent with all of our previous analyses that included the Visual-only condition. One of the 2 subgroups showed an edge between the motor and auditory areas, and this subgroup was populated more with the condition that required a motor response (of the 34 runs sorted into this subgroup, 21 came from the Face-Motor-Auditory condition). However, the Visual-only condition was again split between subgroups. This will be discussed below and highlights the potential dangers of using passive tasks to differentiate groups. This is a 63% classification accuracy in clustering individuals with others who are performing the same task. Lastly, although previous Monte Carlo simulations suggest S-GIMME's performance is not dependent on the number of subgroups (Gates et al., 2017), we note in these data S-GIMME identified the incorrect number of subgroups whenever the expected number of subgroups was not two. Future work should continue to investigate the effect of the number of subgroups on S-GIMME performance.

For comparison purposes, we considered two popular clustering methods: Gaussian finite mixture modeling (McLachlan & Chang, 2004) and the k-means algorithm (MacQueen, 1967). For each of the comparison approaches, individual-level multivariate time series were standardized and then summarized by their (vectorized) first-order VAR transition matrices. This dimension reduction provided a 60×169 matrix of estimated coefficients to be used as input for the subgrouping procedures. Across both approaches we also considered up to 10 potential subgroups and the best solution was selected according to an established external criterion.

In the first approach, we used the vectorized individual-level VAR(1) estimates as input for Gaussian finite mixture modeling (McLachlan & Chang, 2004) to identify subgroups from the time series dynamics. To accomplish this, we used the `mclust` R package (McLachlan

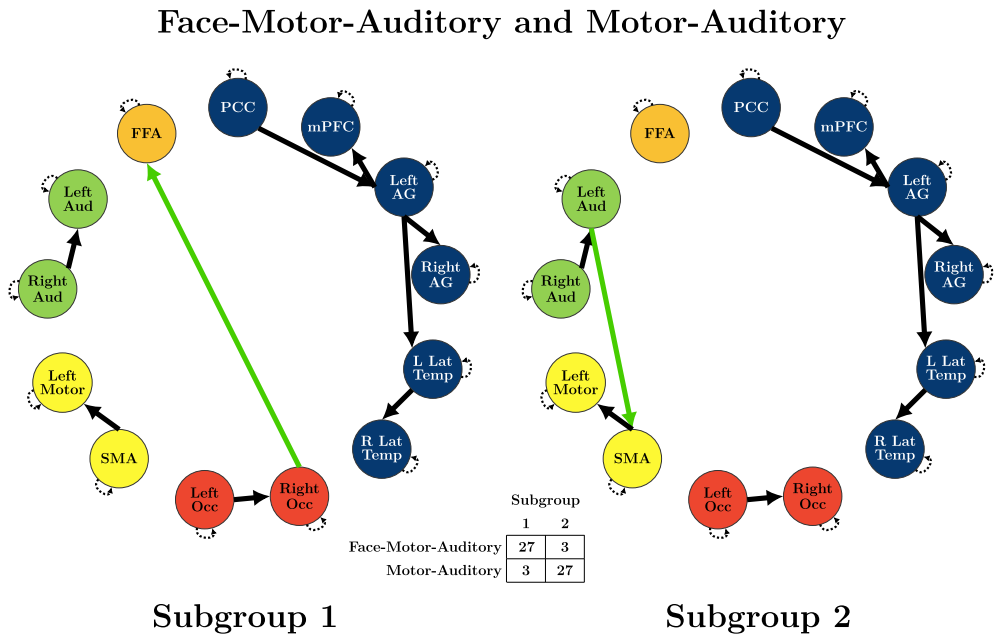

FIGURE 6.
S-GIMME results for 2 tasks: Face-Motor-Auditory and Motor-Auditory. Thick black lines represent edges that were consistent across all subjects (significant for at least 75%). Green lines represent subgroup-level edges (edges that were significant for at least 75% of subjects within a subgroup). Regions labeled and defined as in Table 1 and Fig. 2.

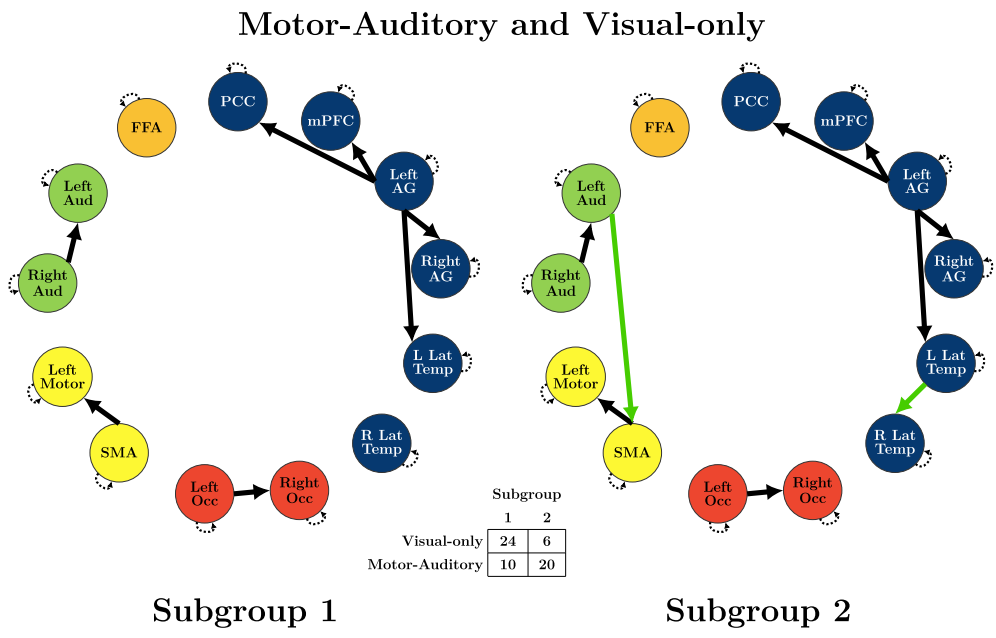

FIGURE 7.
S-GIMME results for 2 tasks: Motor-Auditory vs. Visual-only. Thick black lines represent edges that were consistent across all subjects (significant for at least 75%). Green lines represent subgroup-level edges (edges that were significant for at least 75% of subjects within a subgroup). Regions labeled and defined as in Table 1 and Fig. 2.

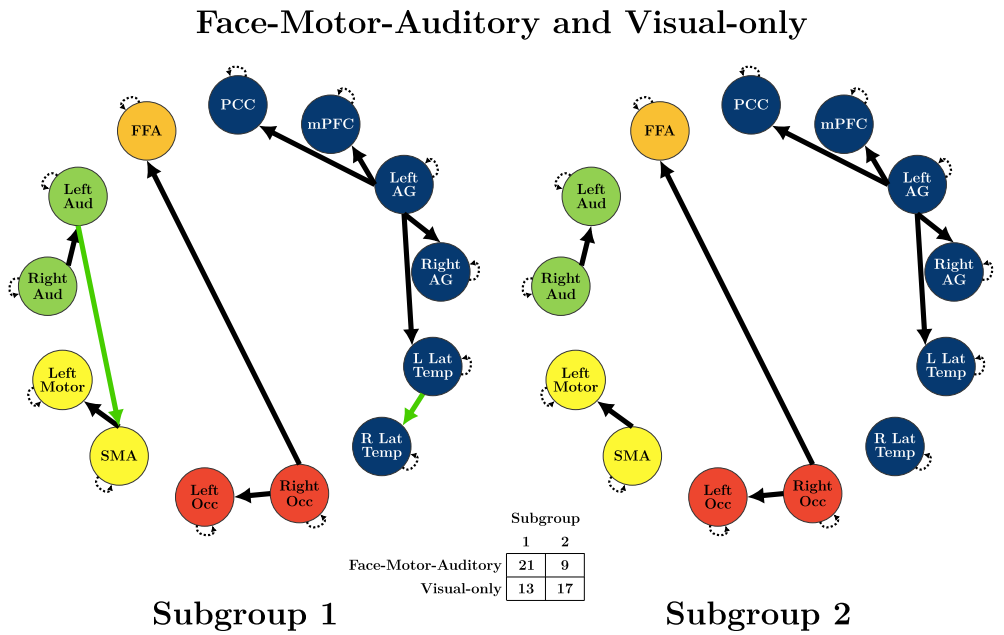

FIGURE 8.
S-GIMME results for 2 tasks: Face-Motor-Auditory vs Visual-only. Thick black lines represent edges that were consistent across all subjects (significant for at least 75%). Green lines represent subgroup-level edges (edges that were significant for at least 75% of subjects within a subgroup). Regions labeled and defined as in Table 1 and Fig. 2.

& Chang, 2004). The best fitting subgroup solution, both in terms of the number of subgroups and the optimal covariance parameterization (the volume, shape, and orientation of the covariance constraints across groups), was chosen based on the Bayesian information criterion (BIC; Schwarz, 1978). Across all comparisons, a single subgroup was identified as optimal based on the BIC. Here, we report metrics for the two-group solution, as it was considered the second-most optimal solution, also across all comparison conditions. For the two-group solution, the mixture model approach classified tasks accurately 50% in the Face-Motor-Auditory vs Visual-only task, 55% in the Motor-Auditory vs Visual-only comparison, and 52% of the time in the Motor-Auditory vs Face-Motor-Auditory comparison.

In the second approach, we used the k-means algorithm on the same 60×169 matrix of estimated coefficients for clustering on the time series dynamics. Again we considered up to 10 potential subgroups and identified the optimal number of subgroups using the Calinski-Harabasz index (Cal'inski & Harabasz, 1974) and the Duda-Hart test (Duda et al., 1973). The Duda-Hart test (Duda et al., 1973) was used to determine if there should be more than 1 cluster, and the Harabasz index was used to rank the different cluster solutions. For these analyses, we used the `fpc` package (Hennig, 2020). Unlike the mixture approach, the k-means approach identified two subgroups as optimal across all three comparison. The k-means approach classified tasks accurately 53% of the time in the Face-Motor-Auditory vs Visual-only comparison; 67% of the time in the Motor-Auditory vs Visual-only comparisons; and 57% of the time in the Motor-Auditory vs Face-Motor-Auditory comparison.

The goal of this paper is to document attempts to validate the S-GIMME algorithm based on brain states explicitly induced through different tasks in a new fMRI dataset. For this reason, any comparison with alternative approaches should not be treated as exhaustive. There are many decision points required for conducting unsupervised classification of multivariate time series arising

from multiple subjects. For example, some form of dimension reduction is often required. Here, we chose to use the VAR(1) model for summarizing the individual-level dynamics to be used as input for the cluster analyses. Although this approach is common in the literature, it may be sub-optimal for a number of reasons. First, it may not provide enough dimension reduction, resulting in an input matrix with insufficient signal given the sample size. On the other hand, it may also not provide enough information to obtain accurate classifications. S-GIMME, for example, uses both the lagged and contemporaneous dynamics to identify subgroups. We mention these details to highlight these comparison results should not be used as evidence of S-GIMME's superiority over alternative methods. Instead, these results are intended to provide some preliminary context on the subgroup classification aspects of this example.

6. Discussion

This study aimed to test the effectiveness of S-GIMME in identifying and segregating distinct subgroups using relatively small fMRI datasets (runs of under 8 min; 1 run per subject per condition). We used 3 different task runs to induce slightly different network connectivity and employed the S-GIMME algorithm for analysis. The algorithm was able to differentiate network structures across conditions for one of the three tasks and partly for another, even though each individual contributed a run to each condition in the comparisons. Specifically, a subject's runs were not grouped with their own other runs but rather were grouped with other individuals doing the same task. This speaks to the sensitivity of this unsupervised classification approach.

The results across conditions followed expectations. Across all conditions, edges were consistently found among brain regions that are known to have high correlations across time. Subgroup-specific edges were also consistent with the expected differences between tasks, especially with regard to the Motor-Auditory condition and the Face-Motor-Auditory condition. In the analysis including these two conditions, the primary differences in the two subgroups that S-GIMME identified were: (1) the edge between the FFA and the primary visual areas and (2) the edge between the auditory and motor areas. The FFA and the primary visual edge was consistently found only in the subgroup into which the Face-Motor-Auditory runs were most often sorted. In contrast, the subgroup lacking those edges, but still having persistent edges between primary auditory and motor regions contained most of the runs from the Motor-Auditory runs. These results suggest that S-GIMME can resolve, in a data-driven and unsupervised manner, the subtle differences, as well as similarities, between different mental states. Thus, this algorithm holds promise for use in investigating possible differences between clinical populations, as well as identifying potential subgroups within a population.

One of the three conditions here stood out in not eliciting a consistent network structure when compared to the other conditions. The Visual-only task did not cleanly and consistently subgroup together when analyzed along with the other two conditions. For this reason, S-GIMME separated FaceMotor from Motor-Only tasks well above chance, whereas the same separation was not seen in the other two comparisons which included the Visual-Only Task. This could potentially be due to the fact that the Visual-only condition is the least-constrained task condition we employed; this task provided subjects with the most leeway to set their own mental state. Variability in what participants did during this passive run could relate to the other tasks, which included motor-responses and face stimuli. The mental state for some subjects, during this passive condition, may have been to think about the face stimuli or motor-response requirements of the other conditions they knew were upcoming—and this may be why some subjects' visual-only runs showed fewer than expected differences with those other tasks.

The finding of no clear subgroup for the Visual-only condition, combined with consistent subgrouping for the other two more highly constrained tasks, may have potential ramifications

for the use of resting-state scans to characterize different populations. As shown here, data-driven algorithms are sensitive enough to pick out differences in the cognitive state of the individual, so using resting state scans to investigate differences between groups could be confounded by differences in cognitive state. Forcing subjects into a more constrained mental state by use of demanding tasks may thus increase the ability to ascertain differences in network structures across groups, as opposed to those results being skewed by higher variability in what subjects are choosing to do during passive, unconstrained scans.

On the other hand, the variability in the mental state of subjects during an unconstrained run could be viewed as an advantage if it is better able to identify static trait characteristics because network structures will not be overly influenced by a demanding task. Current understanding of resting state conditions is that the internal phenomena varies across participants and as such might elicit baseline brain processes that reveal idiosyncratic patterns (Hurlburt et al., 2015). Prior work has been able to subset individuals based on diagnostic status using only the resting state data as input to S-GIMME (Price et al., 2017b), with another study finding diagnoses-specific edges based on resting state data collected across four diagnostic categories and applied to a confirmatory subgrouping GIMME approach (Henry et al., 2019b). Future research should address the issue of whether task-based networks or resting-state networks or possibly the comparison of the two, are most powerful in being able to detect the type of brain processes that may be most useful in identifying distinct subgroups.

7. Conclusion

The current findings, obtained from a strictly controlled experiment design, provide evidence that the S-GIMME algorithm can arrive at meaningful subgroups that correspond with externally driven classifications. If the dynamic processes differ consistently and markedly between latent subgroups, then the unsupervised classification approach can find both the underlying subgroups as well as the network structures. It must be stressed that the ability to recover subgroups was related to how well-differentiated the individuals were. In one case, no clear subgroup was found, likely because the task did not create expected differences in dynamic processes. This is critical for researchers to consider as they design their studies and run any sort of unsupervised classification: distinct and meaningful subgroups may not exist in all data sets. The differences between the data-driven subgroups can be tasks, as seen here, symptomology, as seen previously (Price et al., 2017a), or any correlate. Thus, subgrouping algorithms hold great promise for helping to classify individuals based on their dynamic processes, in a data-driven, unsupervised manner that can be an important complement to existing methods of evaluation.

References

Aghabozorgi, S., Shirkhorshidi, A. S., & Wah, T. Y. (2015). Time-series clustering: A decade review. *Information Systems, 53*, 16–38.

Arizmendi, C., Gates, K., Fredrickson, B., & Wright, A. (2021). Specifying exogeneity and bilinear effects in data-driven model searches. *Behavior Research Methods, 53*(3), 1276–1288.

Bellman, R. (1966). Dynamic programming. *Science, 153*(3731), 34–37.

Beltz, A. M., & Gates, K. M. (2017). Network mapping with gimme. *Multivariate Behavioral Research, 52*(6), 789–804.
Beltz, A. M., Gates, K. M., Engels, A. S., Molenaar, P. C., Pulido, C., Turrisi, R., Berenbaum, S. A., Gilmore, R. O., & Wilson, S. J. (2013). Changes in alcohol-related brain networks across the first year of college: A prospective pilot study using fMRI effective connectivity mapping. *Addictive Behaviors, 38*(4), 2052–2059.
Beltz, A. M., & Molenaar, P. C. (2016). Dealing with multiple solutions in structural vector autoregressive models. *Multivariate Behavioral Research, 51*(2–3), 357–373.
Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., et al. (2002). Region of interest analysis using an spm toolbox. In *8th international conference on functional mapping of the human brain*, vol. 16, page 497. Sendai.
Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., & Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment, 23*(4), 425–435.
Brodersen, K. H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W. D., Buhmann, J. M., & Stephan, K. E. (2014). Dissecting psychiatric spectrum disorders by generative embedding. *NeuroImage: Clinical, 4*, 98–111.
Brown, T. A. (2006). Confirmatory factor analysis for applied research.
Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186–198.
Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods, 3*(1), 1–27.
Dajani, D. R., Burrows, C. A., Nebel, M. B., Mostofsky, S. H., Gates, K. M., & Uddin, L. Q. (2019). Parsing heterogeneity in autism spectrum disorder and attention-deficit/hyperactivity disorder with individual connectome mapping. *Brain Connectivity, 9*(9), 673–691.
Dickie, E. W., Ameis, S. H., Shahab, S., Calarco, N., Smith, D. E., Miranda, D., Viviano, J. D., & Voineskos, A. N. (2018). Personalized intrinsic network topography mapping and functional connectivity deficits in autism spectrum disorder. *Biological Psychiatry, 84*(4), 278–286.
Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1756), 20170284.
Duda, R. O., Hart, P. E., et al. (1973). *Pattern classification and scene analysis* (Vol. 3). New York: Wiley.
Duffy, K. A., Fisher, Z. F., Arizmendi, C. A., Molenaar, P. C., Hopfinger, J., Cohen, J. R., Beltz, A. M., Lindquist, M. A., Hallquist, M. N., & Gates, K. M. (2021). Detecting task-dependent functional connectivity in group iterative multiple model estimation with person-specific hemodynamic response functions. *Brain Connectivity, 11*(6), 418–429.
Easson, A. K., Fatima, Z., & McIntosh, A. R. (2019). Functional connectivity-based subtypes of individuals with and without autism spectrum disorder. *Network Neuroscience, 3*(2), 344–362.
Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. *Psychological Methods, 6*(4), 352.
Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science, 6*(3), 416–427.
Ernst, A. F., Timmerman, M. E., Jeronimus, B. F., & Albers, C. J. (2021). Insight into individual differences in emotion dynamics with clustering. *Assessment, 28*(4), 1186–1206.
Finn, E. S. & Constable, R. T. (2022). Individual variation in functional brain connectivity: Implications for personalized approaches to psychiatric disease. *Dialogues in Clinical Neuroscience*.
Fisher, A. J., & Boswell, J. F. (2016). Enhancing the personalization of psychotherapy with dynamic assessment and modeling. *Assessment, 23*(4), 496–506.
Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping, 2*(4), 189–210.
Gates, K. M., Fisher, Z. F., & Bollen, K. A. (2020). Latent variable gimme using model implied instrumental variables (MIIVs). *Psychological Methods, 25*(2), 227.
Gates, K. M., Henry, T., Steinley, D., & Fair, D. A. (2016). A monte carlo evaluation of weighted community detection algorithms. *Frontiers in Neuroinformatics, 10*, 45.
Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guiskewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate Behavioral Research, 52*(2), 129–148.
Gates, K. M., & Molenaar, P. C. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage, 63*(1), 310–319.
Gates, K. M., Molenaar, P. C., Iyer, S. P., Nigg, J. T., & Fair, D. A. (2014). Organizing heterogeneous samples using community detection of gimme-derived resting state functional networks. *PLoS ONE, 9*(3), e91322.
Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE, 12*(6), e0174035.
Gumus, M., Mack, M. L., Green, R., Khodadadi, M., Wennberg, R. A., Crawley, A., Colella, B., Tarazi, A., Mikulis, D., & Tator, C. H., et al. (2022). Brain connectivity changes in postconcussion syndrome as the neural substrate of a heterogeneous syndrome. *Brain Connectivity*.
Heller, R., Stanley, D., Yekutieli, D., Rubin, N., & Benjamini, Y. (2006). Cluster-based analysis of FMRI data. *NeuroImage, 33*(2), 599–608.
Hennig, C. (2020). *fpc: Flexible Procedures for clustering*. R package version 2.2-9.
Henry, T. R., Feczko, E., Cordova, M., Earl, E., Williams, S., Nigg, J. T., Fair, D. A., & Gates, K. M. (2019). Comparing directed functional connectivity between groups with confirmatory subgrouping gimme. *NeuroImage, 188*, 642–653.

Henry, T. R., Feczko, E., Cordova, M., Earl, E., Williams, S., Nigg, J. T., Fair, D. A., & Gates, K. M. (2019). Comparing directed functional connectivity between groups with confirmatory subgrouping gimme. *NeuroImage, 188*, 642–653.
Hurlburt, R. T., Alderson-Day, B., Fernyhough, C., & Kühn, S. (2015). What goes on in the resting-state? A qualitative glimpse into resting-state experience in the scanner. *Frontiers in Psychology, 6*, 1535.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*(11), 4302–4311.
Lane, S., Gates, K., Fisher, Z., and Molenaar, P. (2021). Package 'gimme'.
Lane, S. T., & Gates, K. M. (2017). Automated selection of robust individual-level structural equation models for time series data. *Structural Equation Modeling: A Multidisciplinary Journal, 24*(5), 768–782.
Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods, 24*(1), 54.
Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M.-Y., Gilmore, A. W., McDermott, K. B., Nelson, S. M., Dosenbach, N. U., et al. (2015). Functional system and areal organization of a highly sampled individual human brain. *Neuron, 87*(3), 657–670.
Liao, T. W. (2005). Clustering of time series data: A survey. *Pattern Recognition, 38*(11), 1857–1874.
Luo, L., Fisher, Z. F., Arizmendi, C., Molenaar, P., Beltz, A., & Gates, K. M. (2022). Estimating both directed and undirected contemporaneous relations in time series data using hybrid-group iterative multiple model estimation. *Psychological Methods*.
Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. New York: Springer.
MacQueen, J. (1967). Classification and analysis of multivariate observations. In *5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324.
McLachlan, G., & Chang, S. (2004). Mixture modelling for cluster analysis. *Statistical Methods in Medical Research, 13*(5), 347–361.
Miller, M. B., & Van Horn, J. D. (2007). Individual variability in brain activations associated with episodic retrieval: A role for large-scale databases. *International Journal of Psychophysiology, 63*(2), 205–213.
Miranda, L., Paul, R., Pütz, B., Koutsouleris, N., & Müller-Myhsok, B. (2021). Systematic review of functional MRI applications for psychiatric disease subtyping. *Frontiers in Psychiatry*, 12.
Molenaar, P. C. (2017). Equivalent dynamic models. *Multivariate Behavioral Research, 52*(2), 242–258.
Newman, M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E, 69*(6), 066133.
Nichols, T. T., Gates, K. M., Molenaar, P. C., & Wilson, S. J. (2014). Greater bold activity but more efficient connectivity is associated with better cognitive performance within a sample of nicotine-deprived smokers. *Addiction Biology, 19*(5), 931–940.
Olszowy, W., Aston, J., Rua, C., & Williams, G. B. (2019). Accurate autocorrelation modeling substantially improves fMRI reliability. *Nature Communications, 10*(1), 1–11.
Pons, P. & Latapy, M. (2005). Computing communities in large networks using random walks. In *International symposium on computer and information sciences*, pp. 284–293. Springer.
Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., et al. (2011). Functional network organization of the human brain. *Neuron, 72*(4), 665–678.
Price, R. B., Gates, K., Kraynak, T. E., Thase, M. E., & Siegle, G. J. (2017). Data-driven subgroups in depression derived from directed functional connectivity paths at rest. *Neuropsychopharmacology, 42*(13), 2623–2632.
Price, R. B., Lane, S., Gates, K., Kraynak, T. E., Horner, M. S., Thase, M. E., & Siegle, G. J. (2017). Parsing heterogeneity in the brain connectivity of depressed and healthy adults during positive mood. *Biological Psychiatry, 81*(4), 347–357.
Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2016). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience, 19*(1), 165–171.
Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage, 52*(3), 1059–1069.
Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology*, pp. 105–129.
Scherf, K. S., Behrmann, M., Humphreys, K., & Luna, B. (2007). Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Developmental Science, 10*(4), F15–F30.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, pp. 461–464.
Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex, 22*(1), 158–165.
Sörbom, D. (1989). Model modification. *Psychometrika, 54*(3), 371–384.
Sporns, O. (2016). *Networks of the brain*. MIT Press.
Tokuda, T., Yoshimoto, J., Shimizu, Y., Okada, G., Takamura, M., Okamoto, Y., Yamawaki, S., & Doya, K. (2018). Identification of depression subtypes and relevant brain regions using a data-driven approach. *Scientific Reports, 8*(1), 1–13.
Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The nimstim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249.
Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 45*(1), 135–170.

Wang, Y., Tang, S., Zhang, L., Bu, X., Lu, L., Li, H., Gao, Y., Hu, X., Kuang, W., Jia, Z., et al. (2021). Data-driven clustering differentiates subtypes of major depressive disorder with distinct brain connectivity and symptom features. *The British Journal of Psychiatry, 219*(5), 606–613.

Ward, J. H., Jr., & Hook, M. E. (1963). Application of an hierarchical grouping procedure to a problem of grouping profiles. *Educational and Psychological Measurement, 23*(1), 69–81.

Weigard, A., Lane, S., Gates, K., & Beltz, A. (2021). The influence of autoregressive relation strength and search strategy on directionality recovery in group iterative multiple model estimation. *Psychological Methods*.

Wright, A. G., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment, 31*(4), 502.

Xu, P., Huang, R., Wang, J., Van Dam, N. T., Xie, T., Dong, Z., Chen, C., Gu, R., Zang, Y.-F., He, Y., et al. (2014). Different topological organization of human brain functional networks with eyes open versus eyes closed. *NeuroImage, 90*, 246–255.

Yang, Z., Xu, Y., Xu, T., Hoy, C. W., Handwerker, D. A., Chen, G., Northoff, G., Zuo, X.-N., & Bandettini, P. A. (2014). Brain network informed subject community detection in early-onset schizophrenia. *Scientific Reports, 4*(1), 1–12.

Yang, Z., Xu, Y., Xu, T., Hoy, C. W., Handwerker, D. A., Chen, G., Northoff, G., Zuo, X.-N., & Bandettini, P. A. (2014). Brain network informed subject community detection in early-onset schizophrenia. *Scientific Reports, 4*(1), 1–12.

Manuscript Received: 28 JUL 2021
Published Online Date: 9 MAR 2023