Subjective difficulty in a verbal recognition-based memory task: Exploring brain-behaviour relationships at the individual level in healthy young adults

Jason Steffenera,*, Chris Habeckb, Dylan Franklinb, Meghan Laua, Yara Yakouba, Maryse Gadb

aInterdisciplinary School of Health Science, University of Ottawa, 200 Lees, Lees Campus, Office # E250E, Ottawa, ON K1S 5S9, Canada
bCognitive Neuroscience Division, Department of Neurology and Taub Institute for Research on Alzheimer’s Disease and The Aging Brain, Columbia University College of Physicians and Surgeons, New York, New York, United States
cSchool of Psychology, University of Ottawa, Ottawa, ON, Canada

A R T I C L E   I N F O

Keywords:
Brain-behavior
Subjective difficulty
Verbal recognition-based memory
fMRI
Individual-level mapping
Sex differences

A B S T R A C T

The vast majority of fMRI studies of task-related brain activity utilize common levels of task demands and analyses that rely on the central tendencies of the data. This approach does not take into account perceived difficulty nor regional variations in brain activity between people. The results are findings of brain-behavior relationships that weaken as sample sizes increase. Participants of the current study included twenty-six healthy young adults evenly split between the sexes. The current work utilizes five parametrically modulated levels of memory load centered around each individual’s predetermined working memory cognitive capacity. Principal components analyses (PCA) identified the group-level central tendency of the data. After removing the group effect from the data, PCA identified individual-level patterns of brain activity across the five levels of task demands. Expression of the group effect significantly differed between the sexes across all load levels. Expression of the individual level patterns demonstrated a significant load by sex interaction. Furthermore, expressions of the individual maps make better predictors of response time behavior than group-derived maps. We demonstrated that utilization of an individual’s unique pattern of brain activity in response to increasing a task’s perceived difficulty is a better predictor of brain-behavior relationships than study designs and analyses focused on identification of group effects. Furthermore, these methods facilitate exploration into how individual differences in patterns of brain activity relate to individual differences in behavior and cognition.

1. Introduction

The majority of functional brain imaging studies rely on the interpretation of group analyses. Group analyses identify the brain regions where the central tendency of activity is common to the participants of each group within a study. This approach defines individual differences as noise or measurement error around the central tendency. It is becoming recognized that a greater understanding of the between-participant variance is needed to fully understand the functioning brain (Lebrerton et al., 2019). Methods that explore the range, source, and effect of individual differences provide many future directions for brain imaging. It facilitates the testing for factors that impact cognitive strategy selection, how demographic differences and lifetime exposures impact patterns of brain activity, structure-function relationships, and the links between brain activity and behavior (Seghier and Price, 2018). Furthermore, identifying and understanding the range of normal variations in individual patterns of brain activity provides insight into the changes that occur in neurodegenerative and psychiatric illnesses (Franzmeier et al., 2020; Tik et al., 2021).

Group analyses identify brain regions used similarly across individuals. Brain-behavior assessments then relate behavioral variables, e.g., task performance, against the variance around the identified central tendency. The weakness in this approach is quantitatively confirmed with the observation that as the sample size increases, the strength of brain-behavior relationships decreases (Grady et al., 2021). These authors explain this phenomenon with the idea that individuals each utilize unique patterns of brain activity when performing the same task. Therefore, the larger the sample, the larger the number of individual patterns of brain activity for performing a task.

Another explanation for individual variance in brain activity when performing a task is that individuals may experience the same task demands differently (Lebrerton et al., 2019). These authors argue that tasks are often developed to produce robust population effects, not interindividual differences. The use of overarching decisions about task de-
mands for all participants in a sample adds variance to the data. This experimentally induced variability results from individuals each experiencing a task with differing levels of subjective difficulty. The various levels of perceived task difficulty likely result in differential levels of brain activity responses. The authors suggest that range adaptation coding principles overcome the limitations of using the same task for all participants. Range adaptation provides each individual with similar subjective input in matched difficulty.

Another source of individual variance in brain activity is sex and gender. A recent review of structural and functional brain imaging results concluded that the brain is not sexually dimorphic (Elliot et al., 2021). However, they concluded that if sex differences did exist, they are buried within individual differences and not detected with group analyses. This conclusion bolsters Grady et al. (2021) conclusion that individuals utilize unique patterns of brain activity when performing the same task, limiting the sensitivity of group-level analyses (Grady et al., 2021). Therefore, sex differences in brain activity may not be in the strength of activation within a common brain region but in how an individual performs a task and the regions used.

The identification of individualized patterns of task related brain activity is not feasible through the use of univariate statistics which consider each brain location in isolation. This is because the strength of univariate statistics is to identify central tendencies in data and not to capture individuality. Univariate analyses identify the brain locations in common use across members of a group typically in response to task demands which assume matched perceived difficulty across people. The assumption of differences in perceived difficulty may be addressed with task titration. This may be performed with a memory task by providing each participant with different study list lengths (Scarmeas et al., 2003). This unfortunately results in task duration differences between participants. Individual differences in perceived task difficulty can also be addressed post-hoc through adjustment of brain activity measures based on task performance (Schneider-Garces et al., 2010). This approach is limited by the fact that it would be statistically circular to both adjust brain imaging data using behavioral measures and to then use them to predict behavior.

The current work drops two of the main assumptions of group studies. It does not identify a central tendency in areas of brain activity responding to task demands across all individuals in a group. Instead, it identifies individualized patterns of brain activity across multiple levels of task demand. Group level comparisons then utilize these participant-specific patterns of brain activity across task demands. Secondly, identically timed and structured tasks are administered to all participants based on subjective difficulty instead of global decisions like the number of items to remember. Finally, sex is included in all models to identify whether sex differences exist in how an individual’s brain activity responds to task demands.

Multivariate analyses identified the individualized patterns of brain activity (Moeller et al., 1996; Spetsiers and Eidelberg, 2010) while participants were engaged in a delayed match to sample working memory task (Rypma et al., 1999; Sternberg, 1966). Multivariate analyses utilize the entire brain in their analyses, unlike univariate analyses, which focus on one location at a time. Focus on brain-wide patterns, therefore, identifies subtle and consistent differences in brain function and minimizes false positives by minimizing the number of statistical tests. Such analyses allow the identification of patterns of brain activity that are specific to each individual. The expression, or utilization, of these brain activity patterns is then related to a range of cognitive demands. The range of cognitive demand used during fMRI scanning is determined based on each individual’s working memory capacity. Demands are therefore parametrically manipulated at levels such as 75, 100, 125% of an individual’s cognitive capacity.

This work addresses the research question of whether, within a sample of twenty-six young adults, differential patterns of brain activity are utilized in similar or different manners to meet increasing task demands and whether there are differences in utilization between the sexes. The null hypothesis is that individuals use a common set of brain regions similarly as task demands increase for both sexes resulting in similar performance. In addition, this work tests whether an individual’s utilization of participant-specific patterns of brain activity is a better predictor of cognitive performance than an individuals’ utilization of a group-derived pattern of brain activity. We show

2. Methods

2.1. Participants

Participants for this study were recruited from the University of Ottawa. Inclusion criteria required participants to be between 18–30 years old, right-handed, normal or corrected to normal vision, and English as a first language. In addition, eligible participants must be in good self-reported health, have no past incidence of a severe head injury, and not be severely ill or hospitalized within the past six months. Participants were excluded if any neurological disorders were likely to affect cognitive function, the use of any psychoactive drugs, and significant cardiovascular disease or atherosclerosis. The study received ethical approval from the Research Ethics Board (REB) of the University of Ottawa, and all participants signed informed consent forms.

2.2. Behavioral task

The delayed match to sample task used visually presented letter stimuli (Rypma et al., 1999; Sternberg, 1966) with relatively minor adaptations from previous use (Hillary et al., 2003; Steffener et al., 2009; Stern et al., 2008). This experiment is trial-based, where each trial consists of three parts, see Fig. 1. A trial begins with presenting letters in a 3 x 3 grid on the screen for 2.5 s. The participant is to study and remember the letters. After removing letters from the screen, a green crosshair appears for 3.5 s. During this time, the participant is to remember the studied letters. Finally, a single probe letter appears for 2.5 s in the center of the screen. During this time window, participants are to determine whether or not they recognize the probe letter as one they studied for this trial. Responses were recorded via a keyboard press.

The memory demands for this task were manipulated based on the number of letters to study at each trial with nine levels of demand possible. When the load level for a trial is less than nine, asterisks fill the

![Fig. 1](image-url) A single trial of the simultaneous presentation of verbal stimuli in the delayed match to sample experiment. The three panels represent the three phases of the experiment. The stimulus encoding phase presents one to nine letters in white font color; the rehearsal phase presents only a crosshair. The recognition phase presents a single probe letter in blue font color.
Fig. 2. A plot of one individual’s changing task demand. As the participant made correct and incorrect responses, the number of letters presented (the task demands) changed. After the initial increase in task demands, task demands increased after three responses in a row, task demands increased. After one error, task demands decreased. Changes in direction are reversals, as seen circled above. Someone’s cognitive capacity is the average of all task demands at reversal values; this is the dashed line for this example.

grid’s remaining locations. The screen position of stimulus letters was the same for all trials of a particular load level. For example, in Fig. 1, a trial with seven letters is presented. The stimulus letters will always be in these positions for all trials with seven letters. The experiment only uses consonant letters (except “W”) to minimize word-forming in the studied letters. No letters in one trial can contain any of the letters used in the previous trial. Letters are not presented in alphabetical order, and study letters are presented in uppercase while probe letters are in lowercase. The use of a lowercase probe ensures that letters are not being matched based on visual features. The task was implemented and administered using PsychoPy2 (Peirce et al., 2019). All software to deliver this task is publicly available at: https://github.com/NCMlab/CognitiveTasks.

2.3. Cognitive capacity

Cognitive capacity is the level of task demand at which a participant consistently performed at 80% accuracy assessed using an adaptive difficulty version of the task, see Fig. 2. The adaptive difficulty used a three-up, one-down staircase design. The procedure started with the lowest level of task demand, one letter, and initially increased in one letter of difficulty for each correct response. After making the first incorrect response, task demands decreased by one letter. After the initial period, task demands increased by one letter after three correct responses. This three up/one down staircase procedure provides approximately 80% accuracy. Each point where the direction of difficulty changed was considered a reversal. Cognitive capacity is the average task demand across all reversal points. As this measure is an average value it is not limited to integer values of load. This procedure was limited to 7 min or 20 reversals (Karmali et al., 2016).

2.4. Task administration in the MRI

The task was administered in the scanner in a block-based manner with five 56 s blocks of six trials each, alternated with 24 s of fixation on a red crosshair. With an introductory rest block of 20 s, the total task time was seven minutes. There was a three-second countdown within each rest block before the beginning of each task block. Two identical runs were administered within the MRI. Task demands for each participant included load levels starting at one letter and continuing up to a load with the number of letters closest to their cognitive capacity and then one letter higher. Therefore, the load levels for every participant were matched on individually assessed cognitive capacity and not on number of letters. The use of block-based administration controls for individual variability in the hemodynamic response to task demands (Handwerker et al., 2004) resulted from differences in neurovascular coupling or vascular reactivity.

2.5. MRI data collection parameters

All neuroimaging used the 3T Siemens Biograph mMR MR-PET scanner at the Brain Imaging Centre (BIC) at the Royal Ottawa Mental Health Centre (ROMHC). Participants wore protective earplugs during the scans and held a squeeze ball they could activate if they felt uncomfortable and wished to terminate the scan.

2.6. Structural MRI

A T1-weighted multi-echo magnetization prepared rapid acquisition gradient echo (MEMPRAGE) image was acquired sagittally (TR = 2530ms; TE 1/2/3/4 = 1.69/3.55/5.41/7.27ms; flip angle = 7°; 1 mm isotropic resolution; 192 slices; 256 mm field of view, ipat (acceleration) = 2, with 32 ref lines and a non-selective inversion time of 1150 ms and 650 Hz/Px BW). Duration: 6:03 min (van der Kouwe et al., 2008).

2.7. Functional MRI

The following acquisition parameters were used for all task-based data collection. A multi-band accelerated EPI sequence (Moeller et al., 2010) using an acceleration factor of 6, TR = 1110ms, TE = 16.6ms, 52-degree flip angle, phase partial Fourier 6/8, 56 slices collected in an alternating increasing slice order, 2.5 × 2.5 mm in-plane resolution, slice thickness = 2.75 mm, field of view: 220 × 200 mm. Each run is 6:38 min long.

2.8. Pre-processing

All image pre-processing and statistical analyses used SPM12 (Wellcome Department of Cognitive Neurology). For each participant’s EPI dataset, images were temporally shifted to correct for slice acquisition order using the middle slice acquired in the TR as the reference. All EPI images were corrected for motion by realigning to the first volume of the first session. The T1-weighted (structural) image was coregistered to the first EPI volume using mutual information. This coregistered high-resolution image was used to determine the transformation into a standard space defined by the Montreal Neurological Institute (MNI) template brain supplied with SPM12 using the new segment tool (Ashburner and Friston, 2005). This transformation was applied to the EPI data and re-sliced using 4th degree B-spline interpolation to 2 × 2 × 2 mm. Finally, an 8 mm FWHM kernel smoothed each image.

2.9. Participant level time-series analysis

The time series modeling for each participant had five regressors of interest, one for each level of memory load. Each block was modeled as a rectangular epoch of 56 s in duration. All regressors of the time series models were convolved with a standard double-Gamma model of the hemodynamic response function (Glover, 2011, 1999). Masking was explicitly applied using all spatially normalized voxels identified as belonging to the brain. The two sessions were modeled together at the first level statistical modeling phase and combined via contrasts. Five contrasts modeled each level of task demand across the two sessions.

2.10. Group-level analysis

An overview of the first-level univariate, second-level univariate, second-level multivariate, and brain-behavior analyses is demonstrated in Fig. 3 and detailed below.
2.11. Univariate analyses

To provide a benchmark for the subsequent multivariate analyses, a repeated-measures univariate analysis looking at load effects was performed. This second-level group analysis used the sandwich estimator (SwE), which appropriately accounts for the within-subject correlation existing in repeated measures data (Guillaume et al., 2014). The analysis included five images per participant, one for each load level, and tested for an effect of a linear slope across load levels.

2.12. Multivariate analyses

Multivariate analyses using scaled subprofile modeling (Moeller et al., 1987; Spetsieris and Eidelberg, 2010) to identify the principal components of brain activity across the five levels of task demand. This analysis used the Generalized Covariance Analysis toolbox (https://www.nitrc.org/projects/gcva_pca) (Habeck et al., 2005; Habeck and Stern, 2007). The analyses split the data up along orthogonal axes identified in the data. The number of axes is equal to the number of contributing images minus one. Once the orthogonal axes are identified, the data is projected back onto these axes to identify where along these axes each contributing image lies. These values are the subject scaling factors and are analogous to x and y coordinates of a vector in a Cartesian plane.

The first step in the multivariate analyses was the performance of PCA across all images from all participants. Each participant contributed five contrast images, one for each level of task demand. The first principal component eigenimage was then considered the group-level effect. This image was projected back into the participant level data to calculate the expression, or degree of utilization, of the group effect in each image used in the analyses. These expression values are the group subject scaling factors, referred to as the SSF (group). These scaling factors can be considered how much each individual contrast image contributes to the overall group effect.

The group effect was then removed from participant-level data by subtracting the derived group pattern multiplied by the expression of the pattern for each image.

\[ \text{Residualized Image}_{ij} = \text{Image}_{ij} - \text{GroupPC(Image} _{ij}^{T}\text{GroupPC)}^{T} \]

Image refers to the load-related contrast images calculated from the first level GLM analyses with participants indexed by i and load level by j. The number calculated within the parentheses is the SSF for how much the group derived that particular image expresses PC.

Principal component analysis was again applied to the residualized images within each participant. Each of the first eigenimages derived from the five residualized images per person was considered as that participant’s specific pattern of brain activity. The result is an individualized pattern of brain activity independent of group effects. The expression of this pattern within each individual is the participant subject scaling factors, referred to as the SSF (residual).

The stability of each voxel was assessed using 5000 bootstrap resamples, threshold-free cluster enhancement (Smith and Nichols, 2009), and tested with the percentile method for calculating confidence intervals. Regions were identified as being significantly active if they had a Z score magnitude > 2).

2.13. Statistical analyses

Mixed-level modeling tested for load and sex-related effects in task accuracy and response time behavioral measures while controlling cognitive capacity. Mixed models also tested for load and sex effects in brain imaging measures of pattern expression while controlling for cognitive capacity. Mixed level modeling is similar to repeated measures ANOVA except that it has the additional benefit of allowing each participant to have a random intercept and is superior at controlling type I errors (Barr et al., 2013; Judd et al., 2012). The intercept was a random effect across participants, while load, sex, and cognitive capacity were fixed effects. Model estimation used restricted maximum likelihood, and degrees of freedom were estimated using the Satterthwaite
Table 1

| Accuracy proportions | Response Time in sec |
|----------------------|----------------------|
| Mean (std)           | Mean (std)           |
| Relative Load 1      | 0.95 (0.092)         | 0.75 (0.22)         |
| Relative Load 2      | 0.99 (0.033)         | 0.82 (0.30)         |
| Relative Load 3      | 0.91 (0.11)          | 0.97 (0.36)         |
| Relative Load 4      | 0.83 (0.16)          | 1.07 (0.38)         |
| Relative Load 5      | 0.78 (0.12)          | 1.08 (0.33)         |

method (Satterthwaite, 1946). Testing for the significance of the random effect used the likelihood ratio test. A significant result demonstrates significant variability in intercept values across participants. The interclass correlation (ICC) value is reported, which is the proportion of the total variance in the dependent variable that is accounted for by the random intercept of each participant (Nakagawa and Schielzeth, 2010). It is the proportion of variation in the data attributed to between-participant differences. In the context of identifying cross-participant similarities, i.e., group effects, the smaller the value, the better. A value of zero means that the simpler repeated-measures ANOVA would be as appropriate as the more complex mixed-level modeling. Analyses used Jamovi The Jamovi project (2021). jamovi (Version 1.6) (Computer Software). Retrieved from https://www.jamovi.org.

2.14. Robustness statistics

Leave-one-out (LOO) procedures established the robustness of results. These LOO procedures dropped one male-female pair of participants or one load level and repeated all multivariate analyses and mixed-level modeling. Regarding the analysis diagram in Fig. 3, each LOO resampling repeated all steps after the first level modeling. The distributions of the statistical F-scores for the omnibus tests of each factor are compared against the point estimates using the entire sample of data. Mixed-level modeling for robustness statistics was identical to those described above; however, they were implemented in MatLab instead of Jamovi. When leaving out one pair of participants, there are 169 different combinations.

3. Results

The following sections summarize the results regarding the participants, the behavioral data, the imaging results, the load and sex relationships with brain imaging data, and the brain-behavioral relationships.

3.1. Participants

This study included data from a total of twenty-six participants. The mean (standard deviation) age was 22.9 (2.71), with a range of 19–30 years. Self-reported sex was thirteen female and thirteen male, all reporting as cisgender.

3.2. Behavioral analyses

The mean (std) cognitive capacity was 7.19 (1.06) letters ranging from 4.62 to 8.44. Recall that cognitive capacity is an average measure across task demands when the participant’s responses reflect a reversal in demand, i.e. getting easier or harder. The accuracy and response times across load levels are shown in Table 1. Cognitive capacity did not significantly differ between sexes (t(24) = -0.423, p = 0.676, effect size (Cohen’s d) = -0.166; mean (std) females = 7.10 (0.97), males = 7.28 (1.18)). To summarize the following results, load was only a significant main effect when predicting accuracy. The main effects of load and cognitive capacity for response time were significant. Details are described below.

3.3. Accuracy

Predicting accuracy, the main effect of load was significant (F(4, 96) = 16.411, p < 0.001). The interaction was not significant: load by sex (F(4, 96) = 0.571, p = 0.684). The remaining main effects were also not significant: cognitive capacity (F(1,23) = 0.872, p = 0.360), sex (F(1,23) = 0.043, p = 0.837). Repeated differences between levels of task load demonstrate significant differences in accuracy between loads 2 to 3 (t(92) = 2.686, p = 0.009), and loads 3 to 4 (t(92) = 2.685, p = 0.009). The remaining differences were not significant: loads 1 to 2 (t(92) = 1.447, p = 0.151) and loads 4 to 5 (t(92) = 1.647, p = 0.103). The random component of the model (participant, intercept) was not significant (ICC = 0.012, X²(1) = 0.032 p = 0.859).

3.4. Response time

Predicting response time, the main effect of load was significant (F(4, 96) = 20.870, p < 0.0001) as was the main effect of cognitive capacity (F(1, 23) = 7.097, p = 0.0139). The interaction was not significant: load by sex (F(4, 96) = 0.335, p = 0.854). The main effect of sex was not significant (F(1,23) = 3.594, p = 0.071). Repeated differences between levels of task load demonstrate significant differences in response time between loads 2 to 3 (t(92) = -3.277, p = 0.0015), and loads 3 to 4 (t(92) = -2.051, p = 0.043). The remaining differences were not significantly different: loads 1 to 2 (t(92) = -1.619, p = 0.109) and loads 4 to 5 (t(92) = -0.323, p = 0.748). The random component of the model (participant, intercept) was significant (ICC = 0.673, X²(1) = 74.674, p < 0.0001).

3.5. Univariate analysis results

Univariate analyses tested for significant linear increases in brain activity across load levels and are shown in Fig. 4 and Table 2. Group results demonstrated significant activation within the following brain regions: cerebellum, bilateral insula, bilateral caudate, posterior division of the left middle temporal gyrus extending into superior regions, left middle occipital gyrus, bilateral middle to inferior frontal gyri extending through Brodmann areas 44 and 46, bilateral superior parietal regions, and bilateral supplementary motor areas.

Participant level results for linear load effects are also shown in Fig. 4 qualitatively demonstrating the range of brain activation variability between participants.

3.6. Principal component analysis of brain imaging data

Principal component results are presented as group effects, participant-level effects, subject scaling factors, and brain-behavior effects.

3.7. Group effects

Results from applying principal component analysis to all imaging data collapsed across participants, and loads produced a pattern of brain activity representing the activity common to all participants. Out of the 130 images used to derive the group pattern, the first eigenimage accounted for 12.48% of the total variance. Within each individual, the group pattern accounted for a range between approximately zero and 63% of the voxel-wise variance. The within participant mean-variance accounted for by the group effect pattern across participants within each load was: 5.19, 12.30, 21.81, 25.56, and 22.73%.

The group-level pattern of brain activity is shown in Fig. 5. A clusterwise summary of results appears in Table 3. Regions included in the positive direction are the cerebellum, bilateral superior parietal areas (Brodmann area 7), bilateral supplementary motor areas, the left pre-central gyrus, and the middle occipital gyrus. The following regions are identified in the negative direction: right superior medial frontal gyrus.
Table 2
Univariate group analysis imaging results.

| Region                      | Lat | BA | Xmm | Ymm | Zmm | Z  | k  |
|-----------------------------|-----|----|-----|-----|-----|----|----|
| Cerebellum, 6               | R   | -  | 10  | .74 | .22 | 5.80 | 9631|
| Insula                      | R   | 48 | .36 | 18  | 2   | 3.82 | 280 |
| Caudate                     | R   | 25 | 8   | 12  | 8   | 5.10 | 492 |
| Middle Temporal             | L   | 21 | -58 | -28 | 2   | 3.76 | 40  |
| Middle Occipital            | L   | 17 | -18 | -98 | 6   | 3.83 | 96  |
| Superior Frontal            | L   | 48 | -38 | 14  | 26  | 5.39 | 2395|
| Superior Frontal            | R   | 46 | 34  | 56  | 20  | 3.93 | 191 |
| Superior Parietal           | L   | 7  | -37 | -58 | 52  | 5.41 | 1359|
| Middle Frontal              | R   | 44 | 36  | 26  | 32  | 4.16 | 327 |
| Supplementary Motor Area    | R   | 6  | 3   | 4   | 60  | 5.01 | 2074|
| Superior Parietal           | R   | 7  | 21  | -64 | 63  | 5.03 | 1304|
| Supplementary Motor Area    | L   | 6  | 5   | 6   | 56  | 5.57 | 759 |
| Paracentral Lobule          | R   | 4  | 3   | -40 | 70  | 3.42 | 20  |

Note: Lat: Laterality, k: cluster size. Thresholds were |Z| > 3.09, p < 0.001, uncorrected. – refers to locations without labels in the AAL or Brodmans atlases.

Table 3
Group derived pattern of imaging results.

| Region                      | Lat | BA | Xmm | Ymm | Zmm | ClMax | ClSize |
|-----------------------------|-----|----|-----|-----|-----|-------|--------|
| Middle Occipital            | L   | 17 | -14 | -98 | 7   | 3.98  | 791    |
| Cerebellum, 6               | R   | 19 | 33  | -69 | -18 | 2.41  | 89     |
| Superior Parietal           | R   | 18 | 17  | -93 | -16 | 3.80  | 608    |
| Parietal                    | L   | 7  | -28 | -66 | 58  | 18.34 | 1303   |
| Superior Parietal           | R   | 7  | 24  | -61 | 64  | 25.71 | 1192   |
| Supplementary Motor Area    | L   | 6  | -4  | 5   | 62  | 6.28  | 189    |
| Supplementary Motor Area    | R   | 32 | 2   | 12  | 52  | 3.33  | 73     |
| Olfactory                   | R   | 25 | 2   | 24  | -4  | -2.94 | 557    |
| Angular                     | L   | 39 | -45 | -72 | 38  | -3.60 | 281    |
| Parietal                    | L   | 1  | -1  | -58 | 35  | -3.25 | 445    |
| Superior Frontal            | L   | 8  | -16 | 29  | 59  | -3.99 | 489    |
| Superior Medial Frontal     | R   | 9  | 11  | 54  | 40  | -2.30 | 65     |

Note: Lat: Laterality, mid:midline, k: cluster size. Thresholds were |Z| > 2.00. – refers to locations without labels in the AAL or Brodmans atlases.
Fig. 5. The significance map of the principal component derived from the group data collapsing across participant and load levels. Significance assessed at $|Z| > 2.00$ using 5000 bootstrap resamples, threshold-free cluster enhancement, and the percentile method.

Fig. 6. Map of the number of participants whose individual level principal component was significant at each voxel in the two directions. This map is thresholded so that each colored voxel was significant for at least one participant. It is worth pointing out that removing the group effect from the data does not eliminate that voxel from being involved in a participant-level map. Each participant may utilize a region a significant amount above or below the group level. The largest number of participants having significant activation in a voxel is eight, highlighting the large amounts of variability in the spatial distribution of participant-level brain activation.

(Brodmann area 9), left superior frontal gyrus (Brodmann area 8), left angular gyrus, left precuneus, and right olfactory cortex.

3.8. Participant level effects

The group pattern was removed from each of the load-related contrast images from each participant. The PCA was then applied to each participant’s five residualized images. From the five contrast images for each participant, the PCA calculated four eigenimages and four eigenvalues. Analyses focused on the first eigenimage; therefore, each participant has their own unique pattern. To summarize the variance between participants in their maps, Fig. 6 shows a voxel-wise map of counts. This map colors voxels based on the number of participants that significantly expressed that location in their pattern. The range in counts was from zero to eight. Therefore, eight individuals had the highest degree of commonality in any brain region. The greatest overlap between participants was within the bilateral supplementary motor area extending back to the precentral gyrus, right middle frontal gyrus extending back to precentral gyrus, left posterior parietal cortex, and the right middle occipital gyrus.

The variance across contrast load images within a participant, accounted for by the first eigenimage, had a mean (std) across participants of 50.26% (8.82%) and a range of 38.59–76.63%, see Fig. 7. There was no significant relationship between the concentration of variance...
in eigenimage one and cognitive capacity ($r = 0.330, p = 0.100$). The amount of variance accounted for by the first eigenimage did not differ significantly between the sexes ($t(24) = 1.676, p = 0.107$, effect size (Cohen’s $d$) = 0.657; mean (std) females = 53% (11.2%), males = 47.4% (4.5%)).

3.9. Subject scaling factor

After identifying the patterns of brain activity, the subject scaling factor (SSF) is the expression of the pattern for each contrast image or level of task demand. The SSF was calculated for the expression of the group-derived pattern and for each participant’s derived pattern. The SSF for group and participant effects are plotted in Fig. 8 along with mean SSF values. These plots demonstrate the large amounts of variance in how an individual contributes to the group pattern and the different manners in which brain activity responds to increasing task demands.

3.10. Brain-behavior results

Brain-behavior analyses were performed using the SSFs shown in Fig. 8 which includes both those derived from the group level defined pattern of brain activity and the SSF derived from participant level patterns of brain activity after being residualized with respect to the group pattern. Overall fixed effect results are shown in Table 4.

Table 4
F-statistics for fixed effects omnibus tests.

| Outcome Variable | Predictor Variables |
|------------------|---------------------|
|                  | Load | Sex | CognCap | Load: | Sex |
| SSF (group)      | 39.78 | 5.13 | 4.22 | 0.58 |
| SSF(residual)    | 69.03 | 0.11 | 1.50 | 4.88 |

Note: Significant results at alpha < 0.05 are in bold. SSF(gr) and SSF(res) represent the subject scaling factors for the group defined image. SSF(residual) and SSF(res) represent the participant subject scaling factors after residualizing data with respect to the group pattern. CognCap is the cognitive capacity in number of letters, RT is the response time in seconds, Acc is the task accuracy in proportion correct.

3.11. Group level SSF

Predicting expression of the group derived pattern by each individual, the main effect of load was significant ($F(4, 96) = 39.783, p < 0.001$) as was the main effect of sex ($F(1, 23) = 5.130, p = 0.033$), see Fig. 9A. The interaction between load and sex was not significant ($F(4, 96) = 0.580, p = 0.678$). The main effect of cognitive capacity was not significant ($F(1, 23) = 4.216, p = 0.0516$). Repeated differences between levels of task load demonstrate significant differences in expression between loads 1 to 2 ($t(96) = -3.746, p = 0.0003$), and loads 2 to 3 ($t(96) = -5.166, p < 0.0001$). The remaining differences were not significantly different: loads 3 to 4 ($t(96) = -1.186, p = 0.238$) and loads 4 to 5 ($t(96) = 0.287, p = 0.775$). The main effect of sex was driven by greater expression by the males than females ($t(23) = 2.265, p = 0.033$). The random component of the model (participant, intercept) was significant (ICC = 0.548, $X^2(1) = 47.354, p < 0.0001$).

3.12. Participant level SSF after removing group-level effects

Predicting expression of the individually derived patterns after removing group effects, the interaction between load and sex was significant ($F(4, 96) = 4.880, p = 0.0013$) as was the main effect of load ($F(4, 96) = 69.029, p < 0.001$), see Fig. 9B. The main effect of cognitive capacity was not significant ($F(1, 23) = 1.498, p = 0.233$), nor the main effect of sex ($F(1, 23) = 0.114, p = 0.739$). Repeated differences between levels of task load crossed with sex demonstrate signif-
significant differences in expression between loads 3 to 4 (t(96) = -3.132, p = 0.0023). The remaining differences were not significantly different: loads 1 to 2 (t(96) = 1.683, p = 0.100), loads 2 to 3 (t(96) = 1.175, p = 0.243) and loads 4 to 5 (t(96) = -0.671, p = 0.504). The random component of the model (participant, intercept) was significant (ICC = 0.240, $X^2(1) = 9.895$, p = 0.0017).

3.13. Predicting performance with SSFs

The behavioral data analyses presented above are now presented with the inclusion of the expression of the group derived and participant-specific patterns of brain activity to test for brain-behavior relationships. Results demonstrate that the brain measures are unrelated to the measure of task accuracy; however, response time was significantly predicted by expression of the participant-specific patterns of brain activity and not by expression of the group-derived pattern.

3.14. Brain activity and accuracy

Predicting accuracy, the main effect of load was significant (F(4, 102.602) = 7.714, p < 0.001). No interaction effects were significant: load by sex (F(4, 91.878) = 0.674, p = 0.611), load by SSF$_{\text{group}}$ (F(4, 99.979) = 0.300, p = 0.877), and load by SSF$_{\text{residual}}$ (F(4, 105.259) = 0.941, p = 0.443). The remaining main effects were also not significant: cognitive capacity (F(1,26) = 1.138, p = 0.296), sex (F(1,27.142) = 0.014, p = 0.907), SSF$_{\text{group}}$ (F(1, 58.981) = 0.191, p = 0.664), and SSF$_{\text{residual}}$ (F(1, 95.307) = 0.450, p = 0.504). Repeated differences between levels of task load demonstrate significant differences in accuracy between loads 3 and 4 (t(92.887) = 2.537, p = 0.0129). The remaining differences were not significant: loads 1 and 2 (t(107.368) = -0.253, p = 0.801), loads 2 and 3 (t(107.265) = 1.249, p = 0.214), and loads 4 and 5 (t(97.141) = 1.351, p = 0.180). The random component of the model (participant, intercept) was not significant (ICC = 0.025, $X^2(1) = 0.125$ p = 0.724).

3.15. Brain activity and response time

Predicting response time, the interaction between load and SSF$_{\text{residual}}$ was significant (F(4, 88.763) = 3.197, p = 0.017), see Fig. 9C. The main effects of load (F(4, 89.523) = 5.320, p < 0.001) and cognitive capacity (F(1, 23.861) = 7.258, p = 0.013) were both significant. The main effect of cognitive capacity was driven by greater overall expression of brain activity for those with higher cognitive capacity. The interactions
between load and sex (F(4, 86.734) = 0.341, p = 0.850) and load and SSF\textsubscript{group} (F(4, 87.819) = 0.956, p = 0.436) were non-significant. The other main effects were also non-significant: sex (F(1, 24.315) = 2.841, p = 0.105), SSF\textsubscript{group} (F(1, 104.606) = 0.019, p = 0.889), and SSF\textsubscript{residual} (F(1, 95.237) = 0.560, p = 0.441). Repeated differences between levels of task load crossed with SSF\textsubscript{residual} demonstrate a significant difference in response times between loads 3 and 4, reflecting those with the lowest expression of brain activity had large increases in brain activity and response times and those with high overall levels of brain activity were minimally differences between the load levels, (t(87.192) = 3.284, p = 0.001). The remaining differences were not significant: loads 1 and 2 (t(89.409) = -0.198, p = 0.844), loads 2 and 3 (t(88.333) = -0.033, p = 0.973), and loads 4 and 5 (t(89.264) = -1.757, p = 0.082). The random component of the model (participant, intercept) was significant (ICC = 0.692, \(X^2\)\textsubscript{1}(1) = 70.304 p < 0.001).

### 3.16. Robustness results

Leave-one-out (LOO) statistical analyses evaluated the robustness of the current findings when leaving out every possible pair of male and female, 169 possibilities, Table 5 displays the results. The distributions of F-values for all models and their parameters are shown in Supplemental Figs. 1–4. The load effects were significant for all LOO pairs when predicting brain activity. The load by sex interaction was significant 89% of the time. The point estimate for cognitive capacity was not significant; however, this effect is significant in 24% of the LOO scenarios. Predicting behavior load effects was robust in nearly 90% of the LOO scenarios, and the cognitive capacity was robust 98% of the time when predicting response time. The significant load by brain activity when predicting response time was only significant in 68% of the LOO scenarios suggesting this finding requires greater power to be a reliable result.

Leave-one-out analyses also assessed the robustness of the results to the load levels included in the analyses. Results demonstrate which load levels have the greatest effects on the findings. Results are displayed in Supplemental Figs. 5–8 for each parameter in the mixed level model and the required threshold to reach significance. In the model predicting residualized expression of brain activity, SSF(residual), dropping load levels 3 or 5 increased the strength of the load effect in the model. Removing load level 1 decreased the strength of the load effect. When predicting response time, removing load level one increased the strength of the brain-behavior relationship. For the same model, removing load level 5 greatly decreased the relationship between cognitive capacity and response time. The load by brain activity interaction became non-significant when load levels 1, 2, or 4 were dropped from the analyses. When predicting accuracy, the load effect was decreased when dropping load levels 2 or 5 from the analyses.

### 3.17. Robustness to using one PC

The analyses were repeated, leaving out incrementally more PCs to quantify the robustness of these results to only removing the first group-level principal component. Plots of the F-values for the model predicting expression of the participant level brain maps, response time, and accuracy are shown in supplemental Figs. 9–11. Removing additional group-level PCs reduced the strength of the load effect in the models; however, it remained highly significant. The significant cognitive capacity and sex effects were not substantially affected when predicting response time.

### 3.18. Sex differences

Expression of the PCs demonstrated significantly strong sex differences with a main effect of sex for the group level derived PC and a load by sex interaction in the participant level PC. To further explore this finding, the univariate results were split based on sex and reanalyzed to explore this finding further. Two clusters of activation demonstrated significant sex differences in the linear load effect. These results, not
Table 5
Percentage of results that exceeded the point estimate when dropping sex-matched pairs.

| Outcome Variable | Predictor Variables | Load: | Sex | CognCap | Load: | Sex | SSF (group) | 100 | 61 | 24 | 0 | SSF (residual) | 100 | 0 | 0 | 89 |
|------------------|---------------------|-------|-----|---------|-------|-----|-------------|-----|----|----|---|-------------|-----|---|---|----|
| RT               |                     | Load: | Sex | CognCap | SSFgr | SSFres | Load: | Sex | SSFgr | 0 | 0 | SSFres | 0 | 0 | 0 | 68 |

Note: SSF(residual) and SSFres represent the participant subject scaling factors after residualizing data with respect to the group pattern. CognCap is the cognitive capacity in the number of letters; RT is the response time in seconds; Acc is the task accuracy in proportion correct.

Fig. 10. Voxelwise activation for the main effect of sex using the sandwich estimator. Warm colors reflect activation significantly greater in females than males and cool colors reflect activation significantly greater in males than females.

shown, were driven by differential amounts of negative activity as cognitive demands increased. Tests for the main effect of sex demonstrated significant results in both directions, see Fig. 10.

4. Discussion

This study identified distinct patterns of load-related brain activity common to a group of young adults and specific to each participant. The participant-specific patterns of brain activity were all expressed with similar load-related increases despite each involving different brain regions. Therefore, regardless of these spatial differences, the individuals all used their own specific patterns of brain activity in a similar load-related manner. Brain activity increased as task demands increased and either plateaued at an individual's cognitive capacity or continued to increase. Expression of the group pattern demonstrated a similar relationship and plateau with increasing load. In addition, expression of the participant-specific patterns of brain activity were better predictors of response time task performance than the expression of the group-derived pattern.

The voxel-wise analysis demonstrated robust linear increases in brain activity as task demands increased throughout brain regions commonly identified with similar verbal working memory tasks. Despite robust group results, the patterns of linear load-related activation varied greater in regional extent and amplitude across participants. This range included an absence of significantly strong activation to nearly the entire brain reflecting increases in brain activation. The variations also included differences in directions. Some individuals demonstrated activation in both directions, while others only demonstrated significant increases. The widespread interindividual variations are also reflected in the expression of the group and individually derived eigenimages. These SSF values reflect the eigenimages' expression and provide insight into the global patterns of brain activation increases. These results also demonstrate a wide range of individual differences in how brain activation responds to increasing task demands. They also reflect the differences between what is captured by a group map and a personalized map of brain activation.

These qualitative interpretations of load-related expression of the group and participant-specific brain activity patterns are supported by significant quantitative differences. The expression of the group derived pattern differed by sex, with the male participants having consistently greater expression at all levels of task load. This sex finding contradicts a recent review of sexual dimorphisms which found limited brain activation sex differences (Eliot et al., 2021). The current results suggest that sex differences exist; however, the differences become evident when looking across a wide range of task demands and are missed when comparing brain activity at one or two different load levels. It is also possible that the current experimental design of comparing brain activity at matched levels of difficulty instead of task load made sex differences evident. Robustness analyses demonstrated that the load by sex difference found using the participant level maps was robust 89% of the time. Therefore, this finding was not driven by any particularity of the dataset, despite its relatively small size of 26 participants. On the other hand, robustness analyses showed that the main effect of sex when using the group level defined PC was found 61% of the time.

Reinvestigating the voxelwise results split by sex demonstrated minimal sex by load results. Tests of the main effect of sex demonstrated robust sex differences in both directions throughout the brain. Combining these findings, it is evident that there are robust sex-related differences in brain activity in this sample of young adults.

The multivariate analyses provide a single number to characterize brain activation patterns, facilitating the evaluation of brain-behavior relationships. Results demonstrated a trend-level effect between greater expression of the group-derived pattern across load levels and greater cognitive capacity. It is doubtful that this group-level derived brain-behavior effect would be significant with larger samples. Larger samples decrease the strength of brain-behavior relationships because of the multitude of ways individuals utilize their brains to perform cognitive tasks (Grady et al., 2021). The intraclass coefficients (ICC) values from the LME models predicting expression of the group-derived PC was over 0.5. Therefore, the random intercept in the model reflected over 50% of the variance between participants. When using the expressions derived from participant-level PCs, the ICC drops to 0.25. This drop reflects a greater similarity in individual expression of participant-derived maps than in the group-derived map, despite participant-level maps that differ in their spatial composition.
The expression of participant-specific patterns of brain activity demonstrates significant interactions between load and sex and between load and cognitive capacity. The females demonstrated a peak in their expression scores at the third load level and then a decline for load levels at and above their respective cognitive capacities. The males demonstrated a continual increase in expression across all load levels. Greater expression at a higher load level is also related to greater cognitive capacity. Although the three-way interaction between load, sex, and cognitive capacity was not significant, marginal means demonstrate this load by cognitive capacity effect is solely in the females. Therefore, while the males benefit from increased expression of the group-derived pattern, there is the support that females derive greater benefit from the expression of their individualized patterns. Future work needs to confirm this.

The participant-specific brain activity patterns were utilized similarly across load levels despite no restriction that the same brain regions are used. At most, eight people commonly significantly utilized any single voxel. The lack of overlap may not be surprising due to removing the central tendency from the data, which may have captured all common areas of brain activity. However, it is surprising to see the significant load effect in both the expression of the group and participant-derived patterns. Expression of the participant-specific patterns also better predicted response time task performance than the expression of the group pattern.

Response time performance increased as a function of load level while accuracy performance decreased. This result is expected since task demands were chosen such that accuracy would be approximately 80% at load level four and lower above that. Therefore, there is minimal between-participant variance in accuracy across load levels which is confirmed with an ICC of 0.012 in the LME model predicting accuracy. Load-related increases in expression of the participant level pattern, but not the expression of the group pattern, significantly predicted response time when including both in the model. Larger cognitive capacity also predicted greater response time. This observation is due to the nature of the experimental design itself. Each participant’s specific load levels were chosen based on their cognitive capacity. Therefore, someone with a higher cognitive capacity received task demands with higher load levels. Therefore, this finding explains the positive relationship between response time and load level (Sternberg, 2016, 1966).

The observed load-related responses of brain activity fit into current theories of cognitive aging. Neural capacity describes brain activity reaching a maximum level as task demands increase (Stern, 2009; Stern et al., 2005). When brain activity decreases with increasing task demands after reaching a neural capacity, the activity is described as Compensation-Related Utilization of Neural Circuits Hypothesis (CRUNCH) (Reuter-Lorenz and Lustig, 2005). In previous work with the CRUNCH thesis, differences in brain activity as a function of cognitive load were largely attributed to individual differences in working memory span (Schneider-Garcés et al., 2010). These authors demonstrated this through post-hoc adjustment of their imaging data by measures of task accuracy. This approach was unnecessary in the current work since each participant’s brain imaging data was collected across subjective levels of task demand, negating the need for data adjustments.

However, the current results did not support CRUNCH for two possible reasons. One is that the experiment used task demands controlling for perceived difficulty. Therefore, the greatest level of task demand was only one load item above someone’s cognitive capacity. Observations of declines in brain activity may require task demands well above someone’s cognitive capacity. It is also possible that CRUNCH occurs primarily in older adults and not younger adults like the current experiment. It will be interesting to explore the role of the higher-order principal components in young and old adults to identify if patterns of brain activity remain stable across load levels in both young and old adults. The current work assumes they do.

These analyses retained only the first of four PCs from each individual. These primary PCs accounted for between 39 and 77% of the variance in each participant. Secondary functional processes may occur as task demands increase, which the other PCs would capture. Exploring higher PCs has been done with group-level analyses (Stern et al., 2008; Zaranh et al., 2006). Within participants, some may employ alternate strategies (Miller et al., 2012) at different load levels, thereby utilizing different patterns of brain activity depending on the task demands. Future work will explore this avenue of thought. However, unlike Miller et al. (2012), the current methods can use data-driven analyses to identify alternate strategies based on the brain imaging data and not post-hoc behavioral assessments.

This work is novel in that it uses a methodology that relies on an individual’s pattern of brain activity at their subjective level of task demands. It does not simply explore whether an individual’s variance around a group central tendency at common load levels is predictive of their behavior. However, despite individual differences in patterns of brain activity, there are strong commonalities in the individual utilization of respective patterns. These methods meet the recent suggestions that approaches are needed to understand between-participant variance in brain imaging data (Lebreton et al., 2019). Furthermore, these methods provide avenues for investigating individual differences in using neural resources to meet task demands (Seghier and Price, 2018). Finally, the current work provides a means of addressing the concerning observation that unaccounted for individual-level variance in brain activity weakens group analyses of brain-behavior relationships (Grady et al., 2021).

Many questions are addressable with these methods beyond the capabilities of group analyses. One is the exploration of different cognitive strategies. As mentioned above, the current work only investigated the first PC. It is also noted that this first PC accounted for a wide range of variance. Large expressions of higher-order PCs may indicate an individual utilizing multiple distinct patterns of brain activity. These may reflect different cognitive strategies across different levels of task demands. A second question would be an exploration into physiological differences underlying the individuality in patterns of brain activity. Exploring physiological differences could be done using analyses that fuse structural and functional brain measures. Such approaches include multivariate fusion analysis (Sui et al., 2014) or voxel-wise serial univariate mediation analyses (Steffener et al., 2016). Work with older adults experiencing age-related neural changes will explore this possibility. It is also possible that analyses including genetic, developmental, or lifetime exposures (Steffener and Stern, 2012; Stern, 2002) would shed light on the individual differences in patterns of brain activity. Future work with larger samples will explore how load-related differences in pattern expression differ as a function of individual differences, allowing for the development of brain activity profiles.

There are several limitations in the current work and findings. There is a relatively small sample size of twenty-six. However, robustness analyses, a relative novelty in neuroimaging, demonstrate that the reported results are robust to leave-one-out investigations. This approach dropped male-female pairs of individuals due to the observed sex differences in the data.

Robustness analyses explored the role of task demand choice in the experimental design. Each load level was dropped, and analyses repeated. Results demonstrated that each load level in the design had its own contribution to the results. Dropping load level one, recognizing a single letter, decreased the strength of load-related results. This can be interpreted as the pattern of brain activation at the lowest level of task demands reflects generalized brain activation responding to stimuli and responding and not working memory. Dropping some of the higher load levels increased the strength of load-related results. Therefore, larger amounts of interindividual variation at higher task demands may decrease the strength of statistical results. Such robustness explorations are informative for the interpretation of results and experimental design and planning. The final robustness analysis explored the decision to use a single eigenimage to describe brain activity across all individuals. These results demonstrate that the findings are
robust to the removal of more than one eigenimage and support this decision.

The current findings support the recent demonstration that larger samples are not always better (Grady et al., 2021). In addition, the mixed-level statistical modeling used does not provide standardized effect sizes. Unfortunately, due to how these models partition variance (Rights and Sterba, 2019), there is no agreed-upon way for calculating standardized effect sizes. We instead now report the unstandardized effect sizes of the fixed effect in line with general recommendations (Pek and Flora, 2018). Furthermore, this study did not account for differences in the women’s menstrual cycle phase when testing occurred (Dubol et al., 2021).

Assessment of voxelwise significance at the participant level used voxelwise resampling to perform bootstrapping. This approach is suboptimal as compared to the resampling of full images. At the participant level, there are only five images available, thereby making voxelwise resampling the only available method. Despite this limitation, the significance of voxels in participant-level brain maps is only used to qualitatively investigate the number of participants having any voxel significantly active. Furthermore, the threshold-free cluster enhancement method was applied to each bootstrap to minimize the limitations of this approach.

Despite these limitations, there are multiple strengths of the current work. This study used individualized patterns of brain activity derived over a wide range of cognitive demands. Pre-assessment of each individual’s working memory cognitive capacity allowed task demand delivery in the MRI at the same perceived difficulty level for all individuals. Using linear mixed statistical models is also superior at controlling type I errors than alternative repeated measures ANOVA models. Therefore, results from mixed models have a greater likelihood of generalizing to new data sets (Barr et al., 2013; Judd et al., 2012).

5. Conclusion

We demonstrated that utilization of an individual’s unique pattern of brain activity in response to increasing perceived difficulty is a stronger measure of brain-behavior relationships in comparison to study designs and analyses focused on identification of group effects. This approach facilitates exploration into how individual differences in patterns of brain activity relate to individual differences in behavior and cognition. Participants in this sample demonstrated a wide range of distinct patterns of brain activity when performing the same task at matched levels of perceived difficulty. Despite spatial differences in patterns of brain activity, there were substantial similarities in how the expression, or usage, of these patterns changed as task demands increased. The expression of the participant-specific patterns of brain activity were also better predictors of task performance than the expression of a group-derived pattern. Furthermore, the expression of the group and participant-specific patterns differed between the sexes.

It may be time to reassess some of the main assumptions implicit in the field of neuroimaging. The use of central tendencies aims to identify commonality in the brain regions involved in a task and assumes each person uses the same brain regions in the same manner to perform a task. The current results demonstrate similar task-related responses of brain activity but in a wide range of regions. Using the same task demands for all individuals also assumes a similar perception of the task across all participants. However, as the field further explores individual differences in cognitive strategy and physiological variations in brain activity, this assumption also needs reassessment. Finally, sexual dimorphism in brain activity may only be evident when individual differences in patterns of brain activity are incorporated into analyses.

Data availability

The data that support the findings of this study are openly available in the Center for Open Science at https://osf.io/w9vre/.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Credit authorship contribution statement

Jason Steffener: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. Supervision. Chris Habeck: Methodology, Software. Dylan Franklin: Project administration, Investigation, Writing – review & editing. Meghan Lau: Investigation. Yara Yakoub: Investigation. Maryse Gad: Investigation.

Acknowledgment

Thank you to Rahim Ismaili at the BIC for MRI scanning.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119301.

References

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.
Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68. doi:10.1016/j.jml.2012.11.001.
Dubol, M., Eppeffor, C.N., Sacher, J., Fletzer, B., Dernl, B., Lanzenberger, R., Sundström-Poromaa, I., Comasco, E., 2021. Neuroimaging the menstrual cycle: a multimodal systematic review. Front. Neuroendocrinol. 60, 100879.
Elliot, L., Ahmed, A., Khan, H., Patel, J., 2021. Dump the “dimorphism”: comprehensive synthesis of human brain studies reveals few male-female differences beyond size. Neurosci. Biobehav. Rev. 125, 667–697.
Franzmeier, N., Dewenter, A., Fronczkowksi, L., Dicghams, M., Rubinski, A., Neitzel, J., Smith, R., Strandberg, O., Osenkoppelle, R., Kuerg, K., Duering, M., Hansson, O., Ewers, M., 2020. Patient-centered connectivity-based prediction of tau pathology spread in Alzheimer’s disease. Sci. Adv. 6. doi:10.1126/sciadv.abb1327.
Glover, G.H., 2011. Overview of functional magnetic resonance imaging. Neurosurg. Clin. N. Am. 22, 133–139 vii.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9, 416–429.
Grady, C.L., Rieck, J.R., Nichol, D., Rodrigue, K.M., Kennedy, K.M., 2021. Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fMRI data. Hum. Brain Mapp. 42, 204–219.
Guillaume, B., Hua, X., Thompson, P.M., Waldorp, L., Nichols, T.E.Alzheimer’s Disease Neuroimaging Initiative, 2014. Fast and accurate modelling of longitudinal and repeated measures neuroimaging data. Neuroimage 94, 287–302.
Habeck, C., Krakauer, J.W., Ghez, C., Sackheim, H.A., Eidelberg, D., Stern, Y., Moeller, J.R., 2005. A new approach to spatial covariance modeling of functional brain imaging data: ordinal trend analysis. Neural Comput. 17, 1602–1645.
Habeck, C., Stern, Y., 2007. Neural network approaches and their reproducibility in the study of verbal working memory and Alzheimer’s disease. Clin. Neurosci. Res. 6, 381–390.
Handwerker, D.A., Ollinger, J.M., D’Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21, 1639–1651.
Hillary, F.G., Chiaravalloti, N.D., Ricker, J.H., Steffener, J., Bly, B.M., Lange, G., Liu, W.C., Kalnin, A.J., DeLuca, J., 2003. An investigation of working memory rehearsal in multiple sclerosis using fMRI. J. Clin. Exp. Neuropsychol. 25, 956–975.
Judd, C.M., Westfall, J., Kenny, D.A., 2012. Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. J. Pers. Soc. Psychol. 103, 54–69.
Karnali, F., Chaudhuri, S.E., Yi, Y., Merfeld, D.M., 2016. Determining thresholds using adaptive procedures and psychometric ftc: evaluating efficiency using theory, simulations, and human experiments. Exp. Brain Res. 234, 773–789.
Lebreton, M., Buvard, S., Daunizeau, J., Palminteri, S., 2019. Assessing inter-individual differences with task-related functional neuroimaging. Nat. Hum. Behav. 3, 897–905.
Miller, M.B., Donovan, C.L., Bennett, C.M., Aminoff, E.M., Mayer, R.E., 2012. Individual differences in cognitive style and strategy predict similarities in the patterns of brain activity between individuals. Neuroimage 59, 83–93.
Moeller, J.R., Ishikawa, T., Dhawan, V., Spetsieris, P., Mandel, F., Alexander, G.E., Grady, C., Pietrini, P., Eidelberg, D., 1996. The metabolic topography of normal aging. J. Cereb. Blood Flow Metab. 16, 385–398.

Moeller, J.R., Brother, S.C., Sidtz, J.J., Rottenberg, D.A., 1987. Scaled subprofile model: a statistical approach to the analysis of functional patterns in positron emission tomographic data. J. Cereb. Blood Flow Metab. 7, 649–658.

Moeller, S., Yacoub, E., Olman, CA, Auerbach, E, Strupp, J, Harel, N, Ugerbil, K. 2010. Multiband multicoil GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63 (5), 1144–1153.

Nakagawa, S., Schielzeth, H., 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol. Rev. Camb. Philos. Soc. 85, 935–956.

Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindevall, J.K., 2019. PsychoPy2: experiments in behavior made easy. Behav. Res. Methods 51, 195–203.

Pek, J., Flora, DB, 2018. Reporting effect sizes in original psychological research: A discussion and tutorial. Psychol Methods 23 (2), 208–225.

Reuter-Lorenz, P.A., Lustig, C., 2005. Brain aging: reorganizing discoveries about the aging mind. Curr. Opin. Neurobiol. 15, 245–251.

Righini, JD, Sterba, SB, 2019. New Recommendations on the Use of R-Squared Differences in Multilevel Model Comparisons. Multivariate Behavioral Research 1–32.

Rypma, B., Prabhakaran, V., Desmond, J.E., Glover, G.H., Gabrieli, J.D., 1999. Load-dependent roles of frontal brain regions in the maintenance of working memory. Neuroimage 9, 216–226.

Satterthwaite, F.E., 1946. An approximate distribution of estimates of variance components. Biometrics 2, 110–114.

Scarmeas, N., Zarahn, E., Anderson, K.E., Hilton, J., Flynn, J., Van Heernt, R.L., Sackeim, H.A., Stern, V., 2003. Cognitive reserve modulates functional brain responses during memory tasks: a PET study in healthy young and elderly subjects. Neuroimage 19, 1215–1227.

Schneider-Garcés, N.J., Gordon, B.A., Brumback-Pelca, C.R., Shin, E., Lee, Y., Sutton, B.P., Maclin, E.L., Gratton, G., Fabiani, M., 2010. Span, CRUNCH, and beyond: working memory capacity and the aging brain. J. Cogn. Neurosci. 22, 655–669.

Seghier, ML, Price, CJ, 2018. Interpreting and utilising intersubject variability in brain function. Trends Cogn. Sci. 22, 517–530.

Smith, S.M., Nichola, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83–98.

Spetsieris, P.G., Eidelberg, D., 2010. Scaled subprofile modeling of resting state imaging data in Parkinson’s disease: methodological issues. Neuroimage 54, 2899–2914.

Steffener, J., Brickman, A.M., Rakitin, B.C., Gazes, Y., Stern, Y., 2009. The impact of age-related changes on working memory functional activity. Brain Imaging Behav. 3, 142–153.

Steffener, J., Gazes, Y., Habeck, C., Stern, Y., 2016. The indirect effect of age group on switch costs via gray matter volume and task-related brain activity. Front. Aging Neurosci. 8, 162.

Steffener, J., Stern, Y., 2012. Exploring the neural basis of cognitive reserve in aging. Biochim. Biophys. Acta 1822, 467–473.

Sternberg, S., 2016. In defence of high-speed memory scanning. Q. J. Exp. Psychol. 69, 2020–2075.

Sternberg, S., 1966. High-speed scanning in human memory. Science 153, 652–654.

Stern, Y., 2009. Cognitive reserve. Neuropsychologia 47, 2015–2028.

Stern, Y., 2002. What is cognitive reserve? Theory and research application of the reserve concept. J. Int. Neuropsychol. Soc. 8, 448–460.

Stern, V., Habeck, C., Moeller, J., Scarmeas, N., Anderson, K.E., Hilton, H.J., Flynn, J., Sackeim, H., van Heernt, R., 2005. Brain networks associated with cognitive reserve in healthy young and old adults. Cereb. Cortex 15, 394–402.

Stern, Y., Zarahn, E., Habeck, C., Holtzer, R., Rakitin, B.C., Kumar, A., Flynn, J., Steffener, J., Brown, T., 2008. A common neural network for cognitive reserve in verbal and object working memory in young but not old. Cereb. Cortex 18, 959–967.

Sui, J., Hunter, R., Yu, Q., Segall, J.M., Calhoun, V.D., 2014. Function-structure associations of the brain: evidence from multimodal connectivity and covariance studies. Neuroimage 102P1, 11–23.

Tik, N., Livny, A., Gal, S., Gigi, K., Tsafraty, G., Weiner, M., Tavor, I., 2021. Predicting individual variability in task-evoked brain activity in schizophrenia. Hum. Brain Mapp. 42, 2983–3992.

van der Kouwe, A.J.W., Benner, T., Salat, D.H., Fischl, B., 2008. Brain morphometry with multiecho MPRAGE. Neuroimage 40, 559–569.

Zarahn, E., Rakitin, B., Abel, D., Flynn, J., Stern, Y., 2006. Age-related changes in brain activation during a delayed item recognition task. Neurobiol. Aging 28, 784–798.